Earth Syst. Sci. Data, 15, 1005–1035, 2023
https://doi.org/10.5194/essd-15-1005-2023

Open Access Earth System Science Data

Four-century history of land transformation by humans in the United States (1630–2020): annual and 1 km grid data for the HIStory of LAND changes (HISLAND-US)

Xiaoyong Li[1,2,3,4], **Hanqin Tian**[2], **Chaoqun Lu**[5], **and Shufen Pan**[3,2]

[1]State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China
[2]Schiller Institute for Integrated Science and Society, Department of Earth and Environmental Sciences, Boston College, Chestnut Hill, MA 02467, USA
[3]International Center for Climate and Global Change Research, College of Forestry, Wildlife and Environment, Auburn University, Auburn, AL 36849, USA
[4]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China
[5]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011, USA

Correspondence: Hanqin Tian (hanqin.tian@bc.edu)

Received: 20 April 2022 – Discussion started: 18 May 2022
Revised: 22 December 2022 – Accepted: 25 January 2023 – Published: 3 March 2023

Abstract. The land of the conterminous United States (CONUS) has been transformed dramatically by humans over the last four centuries through land clearing, agricultural expansion and intensification, and urban sprawl. High-resolution geospatial data on long-term historical changes in land use and land cover (LULC) across the CONUS are essential for predictive understanding of natural–human interactions and land-based climate solutions for the United States. A few efforts have reconstructed historical changes in cropland and urban extent in the United States since the mid-19th century. However, the long-term trajectories of multiple LULC types with high spatial and temporal resolutions since the colonial era (early 17th century) in the United States are not available yet. By integrating multi-source data, such as high-resolution remote sensing image-based LULC data, model-based LULC products, and historical census data, we reconstructed the history of land use and land cover for the conterminous United States (HISLAND-US) at an annual timescale and 1 km × 1 km spatial resolution in the past 390 years (1630–2020). The results show widespread expansion of cropland and urban land associated with rapid loss of natural vegetation. Croplands are mainly converted from forest, shrub, and grassland, especially in the Great Plains and North Central regions. Forest planting and regeneration accelerated the forest recovery in the Northeast and Southeast since the 1920s. The geospatial and long-term historical LULC data from this study provide critical information for assessing the LULC impacts on regional climate, hydrology, and biogeochemical cycles as well as achieving sustainable use of land in the nation. The datasets are available at https://doi.org/10.5281/zenodo.7055086 (Li et al., 2022).

Published by Copernicus Publications.

1 Introduction

Land use and land cover (LULC) change is an essential component of global change, which has directly transformed ecosystems across most of Earth's ice-free land area (Ellis et al., 2021). The human-induced LULC changes, such as cropland expansion, deforestation, and urbanization, have profound impacts on climate change, carbon and nitrogen cycles, and biodiversity (Houghton et al., 1999; Dangal et al., 2014; Domke et al., 2020; Lark et al., 2020; Tian et al., 2020). On the other hand, sustainably managing agricultural and forest lands has been recognized as a critical pathway to achieve climate mitigation targets (Grassi et al., 2017; Griscom et al., 2017). Thus, a better understanding of historical LULC and its spatial–temporal dynamics is critical for quantifying the effects of LULC change on the ecosystem and climate (Winkler et al., 2021).

In the past four centuries, the conterminous United States (CONUS) has experienced dramatic land use and land cover (LULC) changes associated with land clearing, cropland reclamation, and urban land expansion (Steyaert and Knox, 2008; Drummond and Loveland, 2010; Oswalt et al., 2014; Sohl et al., 2016). Before the arrival of Europeans, indigenous agriculture and crop planting existed in the eastern woodlands, the Great Plains, and the South (Hurt, 2002). Since the first colony in Virginia was established in 1607, cropland and pasture began to expand by land clearing, which initially occurred in the eastern United States. During the colonial era, most people lived in the east of the Appalachian Mountains, and agriculture was the primary livelihood. In the 19th century, territorial expansion (e.g., Louisiana Purchase) opened up new areas for agriculture. Driven by the western movement, land clearing, agriculture expansion, and deforestation expanded across the Appalachian Mountains into Ohio, the Mississippi River basin, and the Great Lakes (Cole et al., 1998; Billington and Ridge, 2001; Steyaert and Knox, 2008; Yu and Lu, 2018). In the Mississippi River valley and Alabama, hardwood forests were cleared for cotton and grain production (Hanberry et al., 2012). The center of lumber production was shifted from the Northeast to the Great Lakes in the 1850s (Schulman, 1973). In California, agriculture and ranching expanded throughout the state, and it soon became an exporter of wheat as the gold mining waned (Olmsted and Rhode, 2017). Entering the 20th century, cropland and pasture in New England, the Atlantic coast, and the Southeast were abandoned (Foster, 1992; Hall et al., 2002; Jeon et al., 2014). The environmental protection movement originated in the 1880s. Both tree planting and forest regeneration from abandoned agricultural land accelerated forest restoration (Stanturf et al., 2014). In the following 90 years, the national total plantation forest area increased to 27 million hectares (Mha) (Oswalt et al., 2014, 2019; Chen et al., 2017). While general trends in historical US landscape change are known, we still lack a long-term and spatially explicit LULC dataset to characterize historical LULC trajectories for the CONUS.

Several efforts have produced LULC data for the CONUS in the past several decades. For example, multiple contemporary and spatially explicit LULC products with a resolution from 30 m to 1 km are available, including Global Land Cover (GLC) 2000 (Bartholome and Belward, 2005), MODIS land cover (Friedl et al., 2010), GlobeLand30 (Chen et al., 2015), National Land Cover Database (NLCD) (Yang et al., 2018; Homer et al., 2020), and Cropland Data Layer (CDL) (Boryan et al., 2011; Lark et al., 2017, 2021). However, these datasets were generated using remote sensing images and cannot be used to characterize the century-long land use dynamics. Global-scale and long-term coverage land use datasets (e.g., Land and Use Harmonization (LUH2), the History Database of Global Environment (HYDE)) are widely used in global climate simulations and carbon budget projects (Klein Goldewijk et al., 2017; Hurtt et al., 2006, 2020). However, these datasets have a coarse resolution (from 5 arcmin to 0.25°), which cannot present regional-scale details well (Li et al., 2016; Yu and Lu, 2018). Moreover, the data uncertainties will significantly impact the quantification of LULC effects on the ecosystem (Peng et al., 2017; Yu et al., 2019). Some studies focused on reconstructing historical single-type land use datasets (e.g., built-up area and cropland) for the US (Zumkehr and Campbell, 2013; Yu and Lu, 2018; Leyk et al., 2020). Nevertheless, the dynamics of pasture, forest, shrub, and grassland also profoundly impact ecosystem carbon dynamics (Chen et al., 2006; Tian et al., 2012). Therefore, developing a long-term and high-resolution LULC dataset with multiple types for the CONUS is essential for understanding the LULC history and LULC impact on ecosystem dynamics, regional climate, hydrology, carbon and nitrogen cycles, and greenhouse gas emissions.

In this study, we aim to reconstruct the HIStory of LAND use and land cover for the conterminous United States (HISLAND-US) and analyze the spatial and temporal patterns of LULC changes during 1630–2020 by integrating high-resolution satellite-based LULC data, reliable inventory data, and model-based LULC data. This study consists of three parts: a description of input data and methods, an analysis of spatiotemporal characteristics of LULC in the past four centuries, and a comparison between our results and other studies. We also discussed the driving forces of LULC changes and the uncertainties of the newly developed dataset.

2 Materials and method

This study reconstructed the LULC history (1630–2020) at an annual time step and 1 km × 1 km spatial resolution for the CONUS (48 states) using remote sensing-based LULC data, model-based land use data, and historical census data. In addition, we aggregated the state-level data into eight subregions to analyze the regional divergence of LULC

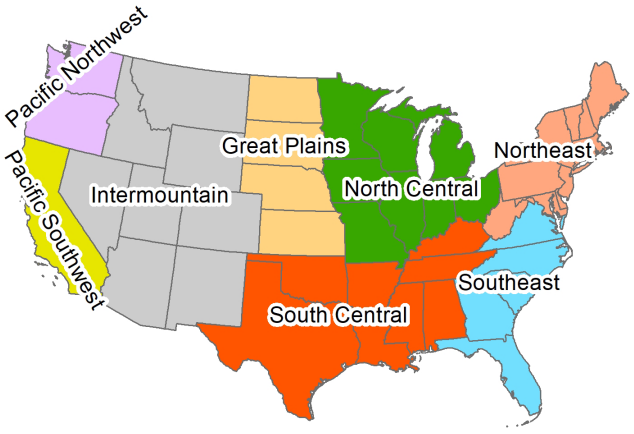

Figure 1. The division of the conterminous United States into eight subregions for data synthesis and analysis in this study.

changes. These subregions include Northeast, North Central, Southeast, South Central, Great Plains, Intermountain, Pacific Northwest, and Pacific Southwest (Oswalt et al., 2014, 2019) (Fig. 1).

The reconstruction process of historical LULC data mainly included two parts: (1) reconstructing the historical urban land, cropland, pasture, and forest area at the state level (Sect. 2.2); and (2) generating 1 km × 1 km spatial resolution gridded LULC data (Sect. 2.3). Figure 2 shows the general workflow for generating historical LULC data. The following sections provide a detailed description of the input data and how we process the data.

2.1 Input datasets for land use and land cover reconstruction

The input datasets included satellite-based LULC data (National Land Cover Database, NLCD), model-based land use datasets (i.e., HYDE3.2 baseline), census and inventory data, and other auxiliary data (Tables 1, S1–S4). The spatial data were resampled or aggregated to 1 km × 1 km resolution for further processing. We also collected some other LULC datasets to validate the newly developed dataset, including cropland density (Yu and Lu, 2017), historical fractional cropland areas (Zumkehr and Campbell, 2013), Economic Research Service (ERS) Major Land Uses data (Bigelow and Borchers, 2017), Land Use Harmonization (LUH2) (Hurtt et al., 2020), CONUS historical land use and land cover (Sohl et al., 2016), county-level crops area (Crossley, 2020), and hay area (Haines et al., 2018) (Table A2).

2.2 Historical land use and land cover area reconstruction

2.2.1 Urban land

In this study, we used the same definition for the developed land as NLCD uses for urban land. The developed land in NLCD includes four components: open space, low intensity developed land, medium intensity developed land, and high intensity developed land (Table 2). We used the NLCD developed land area during 2001–2019 as the urban land area baseline. Before 2001, we applied Historical Settlement Data Compilation for the United States (HISDAC-US) (Leyk et al., 2020; Uhl et al., 2021) as input to reconstruct the historical urban land area. The HISDAC-US built-up areas describes the built environment for most of the CONUS from 1810 to 2015 at 5-year temporal and 250 m spatial resolution using built-up property records, locations, and intensity data (Leyk and Uhl, 2018; Uhl et al., 2021). Here, we assumed that the HISDAC built-up areas data could capture the trend of urban land expansion. Then, the historical urban land can be estimated as follows:

$$\mathrm{HistUrban}_{s,t} = \mathrm{HistUrban}_{s,t+1} \times \frac{\mathrm{HISDAC}_{s,t}}{\mathrm{HISDAC}_{s,t+1}}, \quad (1)$$

where $\mathrm{HistUrban}_{s,t}$ and $\mathrm{HistUrban}_{s,t+1}$ are the reconstructed urban land area of state s in year t and $t+1$, and $\mathrm{HISDAC}_{s,t}$ and $\mathrm{HISDAC}_{s,t+1}$ are the HISDAC built-up area of state s in year t and $t+1$.

There are no census data on urban land area before 1810. Following Liu et al. (2010), we used population to estimate the urban land area by assuming that urban land expanded at the same rate as total population during 1630–1810. The urban land area of each state can be calculated as follows:

$$\mathrm{HistUrban}_{s,t} = \mathrm{HistUrban}_{s,t+1} \times \frac{\mathrm{Pop}_{s,t}}{\mathrm{Pop}_{s,t+1}}, \quad (2)$$

where $\mathrm{HistUrban}_{s,t}$ and $\mathrm{HistUrban}_{s,t+1}$ are the reconstructed urban land area of state s in year t and $t+1$, and $\mathrm{Pop}_{s,t}$ and $\mathrm{Pop}_{s,t+1}$ are the total population of state s in year t and $t+1$.

2.2.2 Cropland

The definition of cropland varies in the existing literature and datasets (Zumkehr and Campbell, 2013; Bigelow and Borchers, 2017; Klein Goldewijk et al., 2017; Homer et al., 2020, Table S5). Cropland, defined by the U.S. Department of Agriculture (USDA) Economic Research Service (ERS), includes five components: cropland harvested, crop failure, cultivated summer fallow, cropland pasture, and idle cropland (Table 2). In this study, we only count the cropland harvested area, which includes row crops and closely sown crops, hay and silage crops, tree fruits, small fruits, berries, tree nuts, vegetables and melons, and miscellaneous other minor crops (https://www.ers.usda.gov/

Table 1. Summary of the input datasets.

Data variables	Time period	Resolution	Data sources
National Land Cover Database	2001, 2003, 2006, 2008, 2011, 2013, 2016, 2019	30 m	Multi-Resolution Land Characteristics Consortium, https://www.mrlc.gov/ (last access: 11 August 2022)
Historical Settlement Data Compilation (HISDAC)	1810–2015	250 m 5-year interval	https://dataverse.harvard.edu/dataverse/hisdacus (last access: 11 August 2022)
ERS major land uses	1910–2020	Nation level Annual	https://www.ers.usda.gov/data-products/major-land-uses/ (last access: 10 July 2022)
CAHA cropland harvested area	1879–2017	State level 4- to 10-year interval	https://agcensus.mannlib.cornell.edu/AgCensus/homepage.do (last access: 10 July 2022)
HYDE3.2 cropland (baseline version)	1600–2017	5 arcmin Annual (2000–2017) 10-year interval (1700–2000) 100-year interval (1600–1700)	https://landuse.sites.uu.nl/datasets/ (last access: 13 February 2023))
NRI pasture area	1982–2017	State level 5-year interval	National Resource Inventory Summary Report, 2017, https://www.nrcs.usda.gov/nri (last access: 13 February 2023).)
HYDE3.2 pasture (baseline version)	1600–2017	5 arcmin Annual (2000–2017) 10-year interval (1700–2000) 100-year interval (1600–1700)	https://landuse.sites.uu.nl/datasets/ (last access: 13 February 2023)
Forest area (USDA)	1630–2017	State level 5- to 18-year interval	Forest Resources of the United States, 2017, https://www.fs.usda.gov/treesearch/pubs/57903 (last access: 14 March 2022)
Forest area (FIA)	1630–2000	State level 10-year interval	Forest Inventory and Analysis, https://www.fia.fs.usda.gov/ (last access: 13 February 2023)
Total population	1630–1999 2000–2020	State level Annual State level Annual	Coulson and Joyce (2003), United States' state-level population estimates: colonization to 1999, https://www.census.gov/en.html (last access: 11 August 2022)
Population density	1790–2010	1 km 10-year interval	Fang and Jawitz (2018) https://doi.org/10.6084/m9.figshare.c.3890191
HYDE3.2 population (baseline version)	1600–2017	5 arcmin Annual (2000–2017) 10-year interval (1700–2000) 100-year interval (1600–1700)	https://landuse.sites.uu.nl/datasets/ (last access: 13 February 2023)
The extent of settled area	1630–present		https://maps.lib.utexas.edu/maps/histus.html (last access: 11 August 2022)

Note: ERS: Economic Research Service, U.S. Department of Agriculture; CAHA: Census of Agriculture Historical Archive; HYDE: History Database of the Global Environment; USDA: United States Department of Agriculture; NRI: National Resource Inventory; and FIA: Forest Inventory and Analysis.

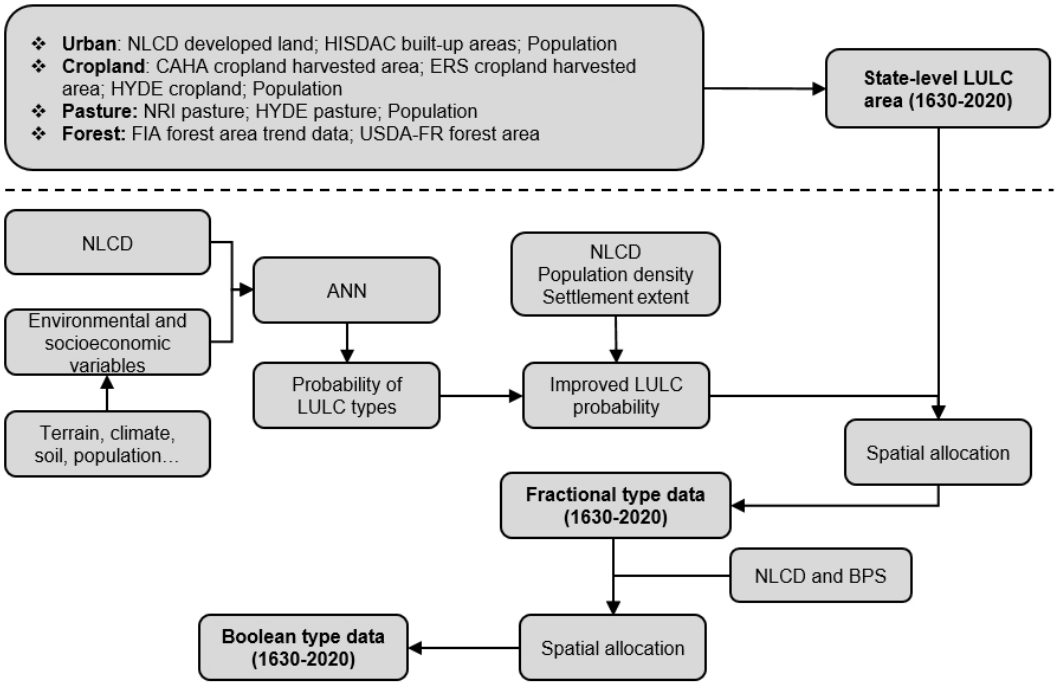

Figure 2. Workflow for generating historical land use and land cover data for the conterminous United States. NLCD: National Land Cover Database; HISDAC: Historical Settlement Data Compilation; CAHA: Census of Agriculture Historical Archive; ERS: Economic Research Service; HYDE: History Database of the Global Environment; NRI: National Resource Inventory; FIA: Forest Inventory and Analysis; USDA-FR: USDA Forest Resources of the United States, 2017; ANN: artificial neural network; and BPS: biophysical settings.

data-products/major-land-uses/glossary/#cropland, last access: 21 August 2022). The USDA Census of Agriculture Historical Archive (CAHA) recorded state-level cropland harvested areas at 4- to 10-year intervals (Tables 1 and S5), which was used for historical cropland area reconstruction between 1879 and 2017. The CAHA cropland was interpolated into annual using the linear method first. To subtract the double-cropped area, we applied the annual national cropland harvested area without double-cropped area from ERS major land uses data to adjust the interpolated cropland harvested area. The adjustment can be expressed as follows:

$$\text{HistCrop}_{s,t} = \frac{\text{cropland harvested}^{\text{linear}}_{s,t}}{\text{cropland harvested}^{\text{linear}}_{\text{conus},t}} \times \text{cropland harvested}^{\text{ERS}}_{\text{conus},t}, \quad (3)$$

where $\text{HistCrop}_{s,t}$ is the reconstructed cropland area of state s in year t, $\text{cropland harvested}^{\text{linear}}_{s,t}$ is the linearly interpolated cropland harvested area of state s in year t based on CAHA cropland harvested area, and $\text{cropland harvested}^{\text{ERS}}_{\text{conus},t}$ is the national total cropland harvested area without double-cropped area in year t. For 2018–2020, the state-level cropland area was calculated based on the state-level area weight in 2017.

For 1879–1910, there was no national-level cropland harvested area without double-cropped area. Therefore, we applied the trend of the CAHA cropland harvested area to reconstruct the historical cropland as follows:

$$\text{HistCrop}_{s,t} = \text{HistCrop}_{s,t+1} \times \frac{\text{CAHA\_CHA}_{s,t}}{\text{CAHA\_CHA}_{s,t+1}}, \quad (4)$$

where $\text{HistCrop}_{s,t}$ and $\text{HistCrop}_{s,t+1}$ are the reconstructed cropland area of state s in year t and $t+1$, and $\text{CAHA\_CHA}_{s,t}$ and $\text{CAHA\_CHA}_{s,t+1}$ are the cropland harvested area of state s in year t and $t+1$.

Because there were no available cropland census data at the state level before 1879, the HYDE cropland was used. We first estimated the cropland per capita by applying the trend of HYDE cropland per capita. Then, the total cropland area can be calculated by multiplying cropland per capita and total population. The data harmonization process can be expressed as follows:

$$\text{HistCrop}_{s,t} = \left(\text{HistCrop\_p}_{s,t+1} \times \frac{\text{HYDE\_Crop\_p}_{s,t}}{\text{HYDE\_Crop\_p}_{s,t+1}}\right) \times \text{Pop}_{s,t}, \quad (5)$$

where $\text{HistCrop}_{s,t}$ is the reconstructed cropland area of state s in year t, $\text{HistCrop\_p}_{s,t+1}$ is the reconstructed cropland per capita of state s in year $t+1$, and $\text{HYDE\_Crop\_p}_{s,t}$ and $\text{HYDE\_Crop\_p}_{s,t+1}$ are HYDE cropland per capita of state s in year t and $t+1$.

2.2.3 Pasture

The definition of pasture also varies among multiple datasets (Klein Goldewijk et al., 2017; U.S. Department of Agriculture, 2020; Table S6). In this study, we use the definition from the National Resource Inventory (NRI), in which pasture is the land that has a vegetation cover of grasses, legumes, and forbs, regardless of whether it is being grazed by livestock or planted for livestock grazing or the production of seed or hay crops (Table 2). The NRI provides a state-level pasture area with a 5-year interval between 1982 and 2017, and we set the pasture area as the baseline for historical reconstruction. Because there were no available pasture census data at the state level before 1982, the HYDE pasture was applied. We first estimated the pasture per capita by applying the trend of HYDE pasture per capita. Then, the total pasture area can be calculated by multiplying pasture per capita and total population. The data harmonization process can be expressed as follows:

$$\mathrm{HistPasture}_{s,t} = \left(\mathrm{HistPasture\_p}_{s,t+1} \times \frac{\mathrm{HYDE\_Pasture\_p}_{s,t}}{\mathrm{HYDE\_Pasture\_p}_{s,t+1}}\right) \times \mathrm{Pop}_{s,t}, \quad (6)$$

where $\mathrm{HistPasture}_{s,t}$ is the reconstructed pasture area of state s in year t, $\mathrm{HistPasture\_p}_{s,t+1}$ is pasture per capita of state s in year $t+1$, and $\mathrm{HYDE\_Pasture\_p}_{s,t}$ and $\mathrm{HYDE\_Pasture\_p}_{s,t+1}$ are the HYDE pasture per capita of state s in year t and $t+1$.

2.2.4 Forest

In this study, we use the forest definition from Forest Inventory and Analysis (FIA), in which forest is defined as land at least 10 % stocked by forest trees of any size, or formerly having such tree cover, with a minimum area classification of 1 acre (Table 2). Two datasets were used for the historical forest area reconstruction. The first is the USDA Forest Resources (USDA-FR) of the United States 2017 (Oswalt et al., 2019). It provides state-level forest areas from 1630 to 2017 with 12 snapshots (i.e., 1907, 1920, 1938, 1953, 1963, 1977, 1987, 1997, 2007, 2012, 2017) and a shot in 1630. Another is FIA's forest area trend data (FATD), which includes state-level forest area from 1760 to 2000 at 10-year intervals and a snapshot in 1630. The data were rebuilt by integrating FIA field data and reports (1950–2000), field inventories (1910–1940), Bureau of the Census land clearing statistics (1850–1900), clearing estimates proportional to population growth (1760–1840), and USDA forest reports. For 1907–2017, the USDA-FR data were used without adjustments. Before 1907, to keep the raw data consistent, we adopted USDA-FR in 1630 as the initial point and gap-fill the missing years by using the changes reflected by FATD data, to reconstruct the forest area between 1630 and 1907. The following harmonization method was conducted to combine the two datasets:

$$\mathrm{HistForest}_{s,t} = \mathrm{USDA\_FR}_{s,1630} \times \frac{\mathrm{FATD}_{s,t}}{\mathrm{FATD}_{s,1630}}, \quad (7)$$

where $\mathrm{HistForest}_{s,t}$ is the reconstructed forest area of state s in year t, $\mathrm{USDAFR}_{s,1630}$ is the USDA–FR forest area of state s in 1630, and $\mathrm{FATD}_{s,t}$ and $\mathrm{FATD}_{s,1630}$ are the FATD forest area of state s in year t and 1630, respectively.

For 2018, 2019, and 2020, we first collected the latest forest area of each state. If a state did not publish the forest area of the latest year, we assumed that the area during these 3 years was the same as that in 2017. The latest forest area data can be accessed at https://fia-usfs.hub.arcgis.com/ (last access: 30 August 2022).

2.2.5 Post-processing of historical urban, cropland, pasture, and forest area

Due to the difference in data sources in the reconstruction step, the total area of urban land, cropland, pasture, and forest may exceed the state's total land area (TLA). Therefore, we calibrated the reconstructed historical land use and land cover area using the following equations:

$$\begin{cases} A^t_{i,\mathrm{rc}}(s) = A^t_{i,\mathrm{r}}(s) & \text{if} \quad \mathrm{TA}^t_{\mathrm{r}}(s) \le \mathrm{TLA}(s) \\ A^t_{i,\mathrm{rc}}(s) = \frac{A^t_{i,\mathrm{r}}(s)}{\mathrm{TA}^t_{\mathrm{r}}(s)} \cdot \mathrm{TLA}(s) & \text{if} \quad \mathrm{TA}^t_{\mathrm{r}}(s) > \mathrm{TLA}(s) \end{cases} \quad (8)$$

$$\mathrm{TA}^t_{\mathrm{r}} = \sum_{i=1}^{n} A^t_{i,\mathrm{r}}(s), \quad (9)$$

where t is the current year; $A^t_{i,\mathrm{rc}}(s)$ and $A^t_{i,\mathrm{r}}(s)$ are the recalibrated area and reconstructed area for the land use class i in the state s, respectively; $\mathrm{TA}^t_{\mathrm{r}}$ is the total area of urban, cropland, pasture, and forest; n is total number of land use types; and s is the state index in the range from 1 to 48.

2.3 Approach for generating gridded land use and land cover data

2.3.1 Calculating the land use and land cover probability

Following previous studies, we applied the "top-down" strategy to allocate the state-level LULC area to the grid level based on probability or suitability surfaces (Fuchs et al., 2013; West et al., 2014; He et al., 2015; Sohl et al., 2016). Previous spatially explicit land use and land cover simulation models, such as the Conversion of Land Use and its Effects (CLUE) model and Forecasting Scenarios of Land use Change (FORE-SCE) model, used the logistic regression (LR) model to develop LULC probability of occurrence (Verburg and Overmars, 2009; Sohl et al., 2014, 2016; Li et al., 2016; Yang et al., 2020). However, it needs to train the LR model for the different units (e.g., county, grid) to calculate a good probability map due to the spatial heterogeneity of land conversion. In comparison, artificial neural networks

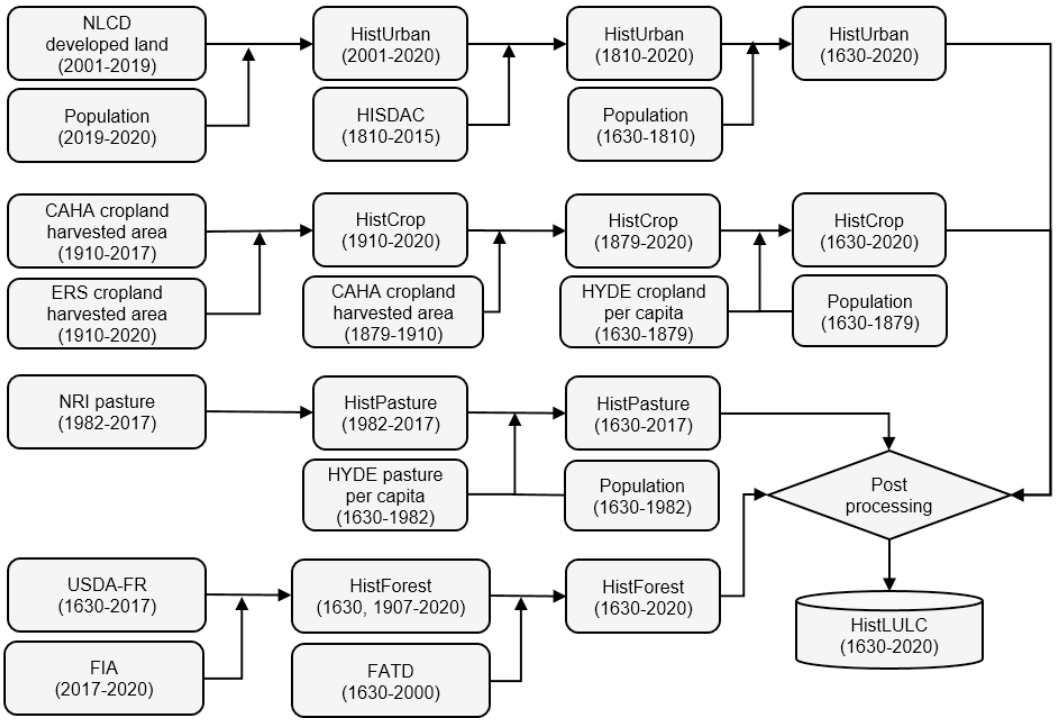

Figure 3. Workflow for reconstructing historical land use and land cover area at the state level. NLCD: National Land Cover Database; HISDAC: Historical Settlement Data Compilation; ERS: Economic Research Service; CAHA: Census of Agriculture Historical Archive; NRI: National Resource Inventory; HYDE: History Database of the Global Environment; FIA: Forest Inventory and Analysis; USDA-FR: USDA Forest Resources of the United States, 2017; and FATD: Forest Area Trend Data. HistUrban, HistCrop, HistPasture, HistForest, and HistLULC refer to historical urban land, historical cropland, historical pasture, historical forest, and historical land use and land cover.

Table 2. Definitions of urban, cropland, pasture, and forest in this study.

LULC	Definition
Urban land	Same as the definition of developed land in the National Land Cover Database (NLCD). Developed land in NLCD includes four components: open space, low intensity developed land, medium intensity developed land, and high intensity developed land (https://www.mrlc.gov/data/legends/national-land-cover-database-class-legend-and-description, last access: 21 August 2022).
Cropland	Same as the definition of cropland in the U.S. Department of Agriculture (USDA) Economic Research Service (ERS) major land use. Cropland defined by USDA ERS includes five components: cropland harvested, crop failure, cultivated summer fallow, cropland pasture, and idle cropland (https://www.ers.usda.gov/data-products/major-land-uses/glossary/#cropland, last access: 21 August 2022). In this study, we only count the cropland harvested area subtracting the double-cropped area.
Pasture	Same as the definition of pasture in the National Resource Inventory (NRI). Pasture is a land cover/use category of land managed primarily for the production of introduced forage plants for livestock grazing.
Forest	Same as the definition of forest from the Forest Inventory and Analysis (FIA). Forest is land at least 10 % stocked by forest trees of any size, or formerly having such tree cover, with a minimum area classification of 1 acre (https://cfpub.epa.gov/roe/definitions.cfm?i=51, last access: 13 February 2023).

(ANNs) can learn and fit complex relationships between input data and training targets and can be used to solve various non-linear geographical problems (Hagenauer and Helbich, 2022). Moreover, ANNs perform better than LR in land use and land cover change simulation (Liu et al., 2017). Therefore, we used the ANN-based Probability of Occurrence Estimation tool in Future Land Use Simulation (FLUS) software to generate the LULC probability (Liu et al., 2017). The independent variables for the ANN model training and prediction include terrain (elevation and slope), climate (annual mean temperature, annual precipitation, annual maximum temperature (July), and annual minimum temperature (January)), crop productivity index, population density, distance to the city, distance to the road, distance to the railway, distance to the river, and soil (soil organic carbon, soil sand, and soil clay) (Table A1). The Boolean type NLCD data in 2001 were used for ANN model training.

Over the past four centuries, the rules of LULC probability change a lot due to the interaction between the natural environment and socioeconomic factors. The contemporary pattern of LULC probability is not representative for the early period (Sohl et al., 2016). Following Klein Goldewijk et al. (2017), we improved the LULC probability by combining the biophysical probability, contemporary probability, population density, human settlement extent, and satellite data. The total probability for each grid cell can be expressed as follows:

$$\begin{cases} \mathrm{TP}_t = (S_{\mathrm{hist}} \times w_1 + S_{\mathrm{satellite}} \times w_2) \times (1.0 + r) & t \le 2001 \\ \mathrm{TP}_t = (S_{\mathrm{satellite}} + \mathrm{Frac\_d}t_{\mathrm{satellite}}) \times (1.0 + r) & t > 2001 \end{cases} \quad (10)$$

$$\begin{cases} \mathrm{Prob}_{\mathrm{hist}} = \mathrm{Prob}_{\mathrm{bio}} \cdot \mathrm{popd}_t \cdot \mathrm{SE}_{\mathrm{weight},t} \\ \mathrm{Prob}_{\mathrm{satellite}} = \mathrm{Prob}_{2001} \cdot \mathrm{SE}_{\mathrm{weight},t} \end{cases} \quad (11)$$

$$\mathrm{SE}_{\mathrm{weight},t} = w_{t0} \times \mathrm{SE}_{t0} + w_{t1} \times \mathrm{SE}_{t1}, \quad (12)$$

where S_{hist} and $S_{\mathrm{satellite}}$ is the LULC fraction generated by using the historical ($\mathrm{Prob}_{\mathrm{hist}}$) and satellite ($\mathrm{Prob}_{\mathrm{satellite}}$) probability; w_1 and w_2 is probability weight; w_1 is set to 0 in 2001 and 100 % in 1850 (and the pre-1850 period as well), while w_2 is set to 0 in 1850 (and the pre-1850 period as well) and 100 % in 2001; $\mathrm{Frac\_d}t_{\mathrm{satellite}}$ is the NLCD LULC fraction dynamics between year t and 2001; $\mathrm{SE}_{\mathrm{weight},t}$ is settlement weight in year t, which is calculated based on the settlement in year $t0$ and year $t1$; and r is a random item with a range of [0, 0.5]. $\mathrm{Prob}_{\mathrm{bio}}$ is the LULC probability that only uses biophysical variables (terrain, climate, and soil variables), Prob_{2001} is the LULC probability that uses all the variables, and popd_t is population density (Fig. S5).

2.3.2 Strategies to generate fractional and Boolean land use and land cover data

Two types of gridded LULC data with 1 km × 1 km spatial resolution were generated. The first is fractional type, in which the dataset includes four fractional components: urban, cropland, pasture, and forest. Another is Boolean type with nine LULC types: urban, cropland, pasture, forest, shrub, grassland, wetland, water, and barren.

To generate the fractional gridded LULC data, we assumed that the fraction of each LULC type at the grid level was determined by the total probability (Fuchs et al., 2013; Tian et al., 2014; West et al., 2014; He et al., 2015). It means that a grid cell (LULC type k) with a high probability will have a high fraction. Based on this principle and the state-level LULC area, we generated the fractional LULC data at 1 km × 1 km resolution and an annual timescale. The detailed information for generating fractional LULC data is shown in the following steps (Fig. 4): (1) prepare the input data – state-level historical LULC area and probability; (2) calculate the state target LULC fraction for type k and initialize an empty LULC fraction map; (3) calculate a temporal fraction layer; (4) modifying the temporal fraction, we assume that the fraction of water and barren is stable, and the sum of urban, crop, pasture, and forest fraction is lower than the maximum fraction in each grid cell; (5) add the temporal fraction data to the empty LULC fraction map; and (6) judge whether the unallocated LULC area is smaller than 0.01 km^2. If yes, the iteration will stop and begin to allocate another LULC type, or else the unallocated area will be assigned to target fraction and return to step (3). The allocation was processed until the unallocated area was less than the threshold (0.01 km^2). The above steps will be conducted for each state and then output the fractional map.

Based on the LULC fraction map, we generated the Boolean type LULC data at 1 km × 1 km resolution. The detailed information is shown in the following steps (Fig. 4): (1) prepare the input data – state-level historical LULC area and LULC fraction data. (2) Generate a temporal LULC map (HistB) through identifying the dominant LULC type in each grid cell and initialize an empty LULC map ($\mathrm{HistB}_{\mathrm{E}}$). (3) Calculate the area difference for LULC type k between the HistB map and target area. (4) If the area difference is negative, we first sort the LULC fraction data where HistB equals to k; the top m (equals to the target area) grid cells where $\mathrm{HistB}_{\mathrm{E}}$ is not assigned a value will be assigned as k; then, if the available number of grid cells (type k) is less than the target area, we will sort the LULC fraction data where HistB map is not equal to k; and the top n (equals to the unallocated area) grid cells where $\mathrm{HistB}_{\mathrm{E}}$ is not assigned a value will be assigned as k. (5) If the area difference is positive, the grid cells where HistB data equal k will be assigned k and $\mathrm{HistB}_{\mathrm{E}}$ not assigned a LULC type will be assigned as k, then we will sort the LULC fraction data where HistB data are not equal to k, and the top n (equals to the unallocated area) grid cells where $\mathrm{HistB}_{\mathrm{E}}$ is not assigned a value will be assigned as k. If steps (4) and (5) finish, the next LULC type will begin to allocate. After the four LULC types of allocation finish, the grid cell that is not assigned a type will be updated using the HistB data and LANDFIRE biophysical settings data (Fig. A1; Rollins, 2009).

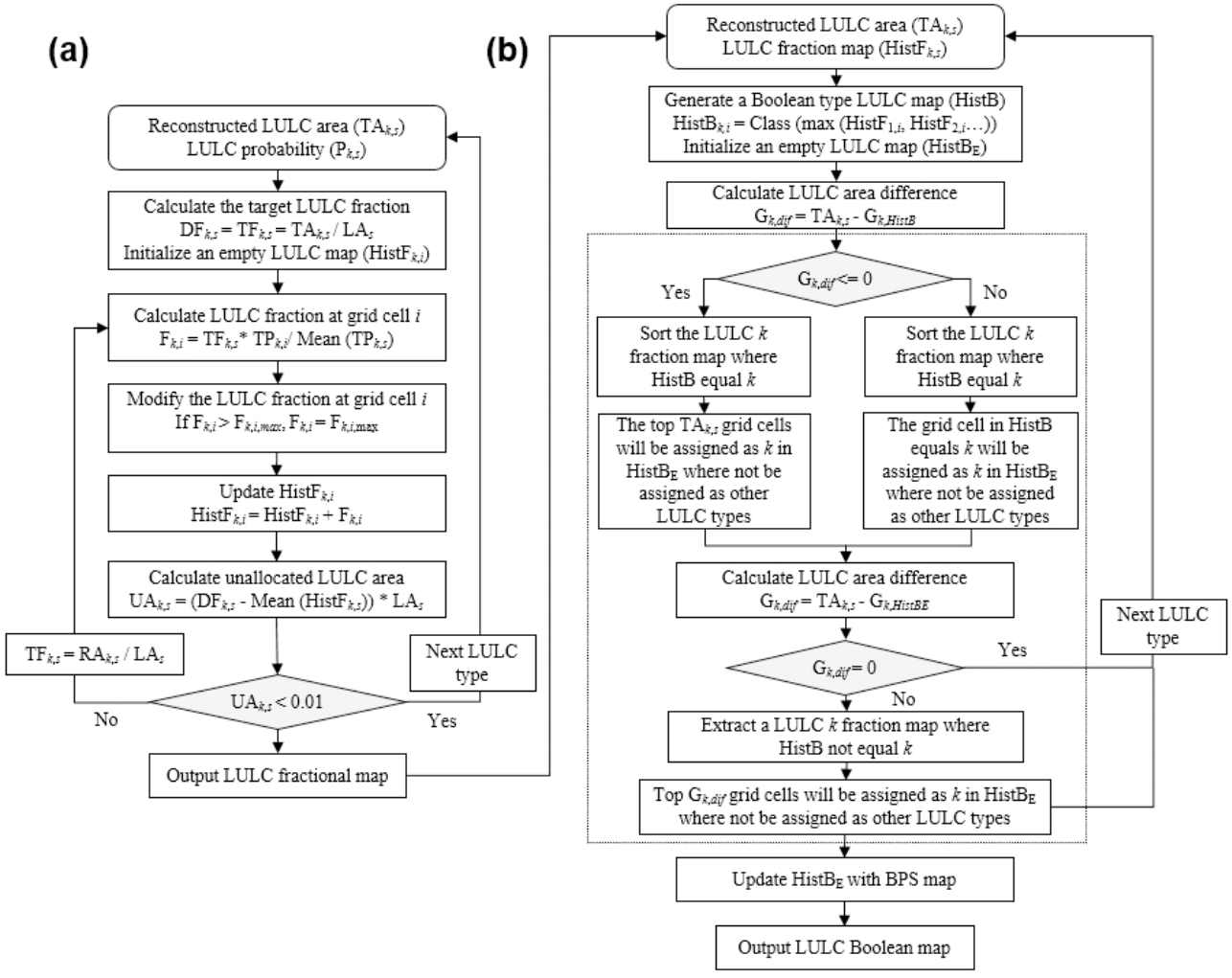

Figure 4. Workflow for generating fractional **(a)** and Boolean **(b)** type LULC data.

2.4 Data comparison

The lack of actual spatially explicit reference data made a complete formal validation impractical. Though the LULC definitions in this study are different from other LULC datasets, data comparison is a way to assess the accuracy of the reconstructed LULC area and spatial pattern. Thus, we conducted three data comparisons to increase the confidence of the newly developed LULC datasets. First, the state-level LULC area derived from the multisource datasets was used for comparison. Considering the differences in the cover period of multiple LULC datasets, we derived the average state-level statistics area for urban, cropland, pasture, and forest from 2000–2020 for comparison. Second, we collected the USDA county-level cropland area between 1840 and 2017 and compared the cropland proportion with that derived from our data in 4 selected years (1850, 1920, 1959, and 2002). Third, we compared urban, cropland, pasture, and forest from the newly developed LULC dataset with the NLCD during 2001–2019 at the grid level.

3 Results

3.1 Land use and land cover change during 1630–2020 in CONUS

The results showed that the LULC change from 1630 to 2020 was characterized by the expansion of cropland and urban land and the shrinking of natural land cover (e.g., forest, grassland, and shrub) (Figs. 5, A2–A5). In 1630, the primary landscape was the forest in the eastern US and Pacific Coast, grassland in the Great Plains, and shrub in the Rocky Mountains (Fig. 5). Urban land, cropland, and pasture were mainly distributed in the east of the US before 1850. Rapid cropland and pasture expansion occurred in the North Central (e.g., Iowa, Illinois, Minnesota), the Great Plains, and the Mississippi River valley during 1850–1920 (Figs. 5 and A3).

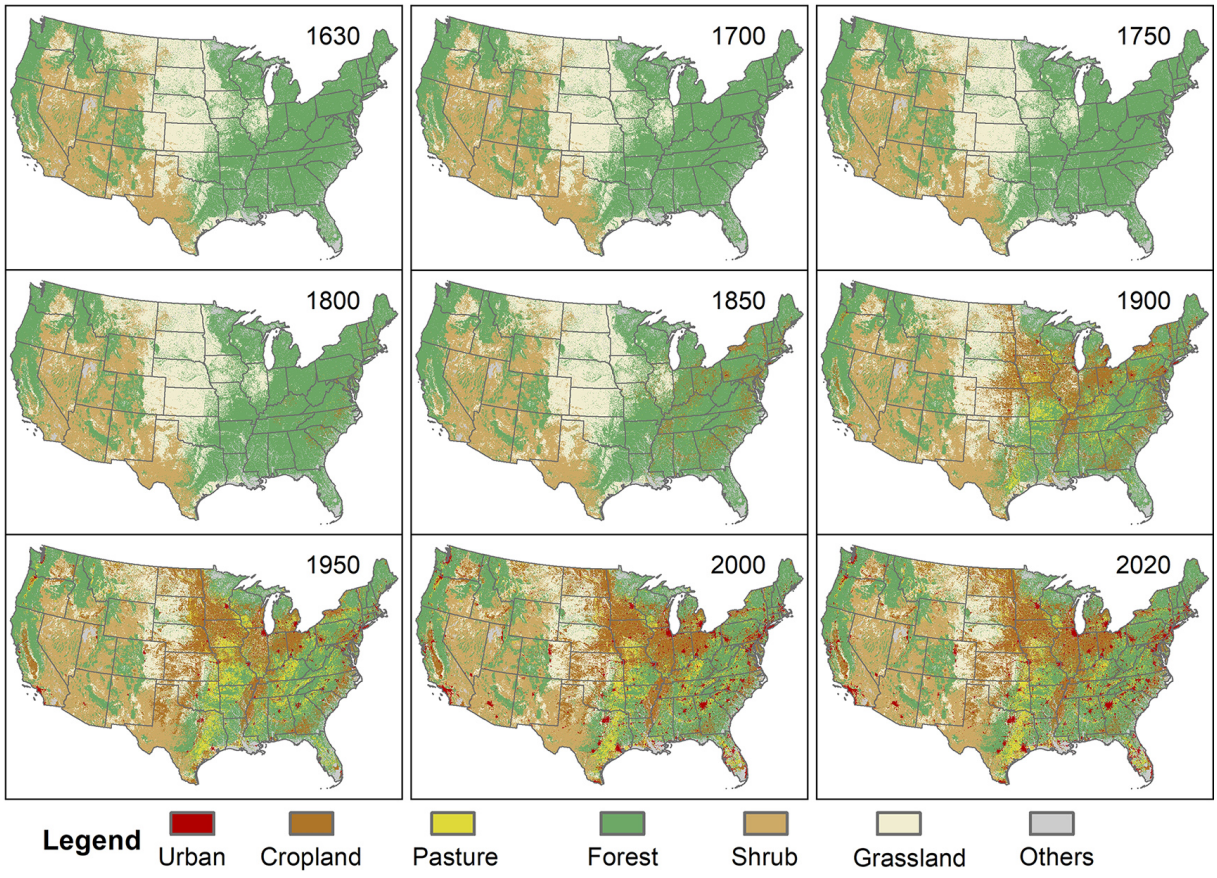

Figure 5. Spatiotemporal patterns of land use and land cover in the conterminous United States during 1630–2020.

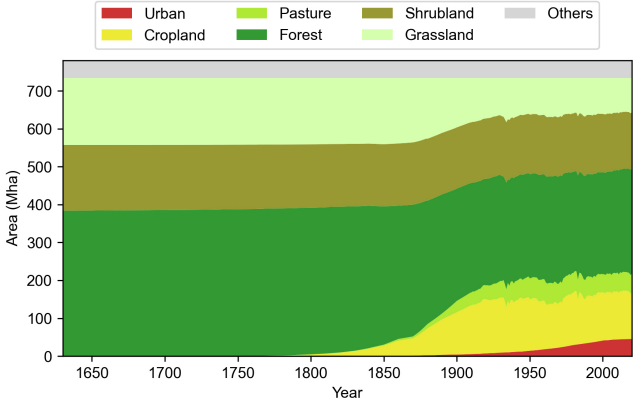

Figure 6. Land use and land cover changes in the conterminous United States from 1630 to 2020.

After 1920, the distribution of major LULC types became relatively stable (Fig. 6). In the 2000s, the cropland in the Corn Belt regions, Central California, and Mississippi Alluvial Plain had the highest cropland density (Fig. A3), and the highest pasture density was found in the east of Texas, Oklahoma, Missouri, and Kentucky (Fig. A4).

The US experienced the colonial era, the war of independence, and territorial expansion between 1630 and 1850. In this period, urban land increased by 0.80 Mha with a total population growth of 23 million (Fig. 6). In the mid-1800s, the cheap land and the industrial revolution attracted many European and Mexican immigrants, which accelerated urban development. In the second half of the 19th century, the population tripled, and the total urban land increased to 4.3 Mha in 1900 (Fig. 6). During the 20th century, both the rapid growth of population and urban land per capita accelerated urban land expansion. Our results show that the urban land per capita increased from 0.02 ha/person in 1900 to 0.14 ha/person in 2020 (Fig. S7). The national total urban land area increased to 45.46 Mha in 2020.

Cropland expanded slowly by 27.09 Mha from 1630 to 1850, and it increased substantially to 142.05 Mha in the following 70 years (Fig. 6). Agriculture intensified after 1920, but the total cropland area in the CONUS was relatively constant, with a peak area of 146.08 Mha in 1932 (Fig. 6). Due to the competition with the high production land in the Midwest, cropland abandonment occurred in the northeast, south, and southeast (Bigelow and Borchers, 2017; Yu and Lu, 2018). During 1950–1975, the rise of the manufacturing

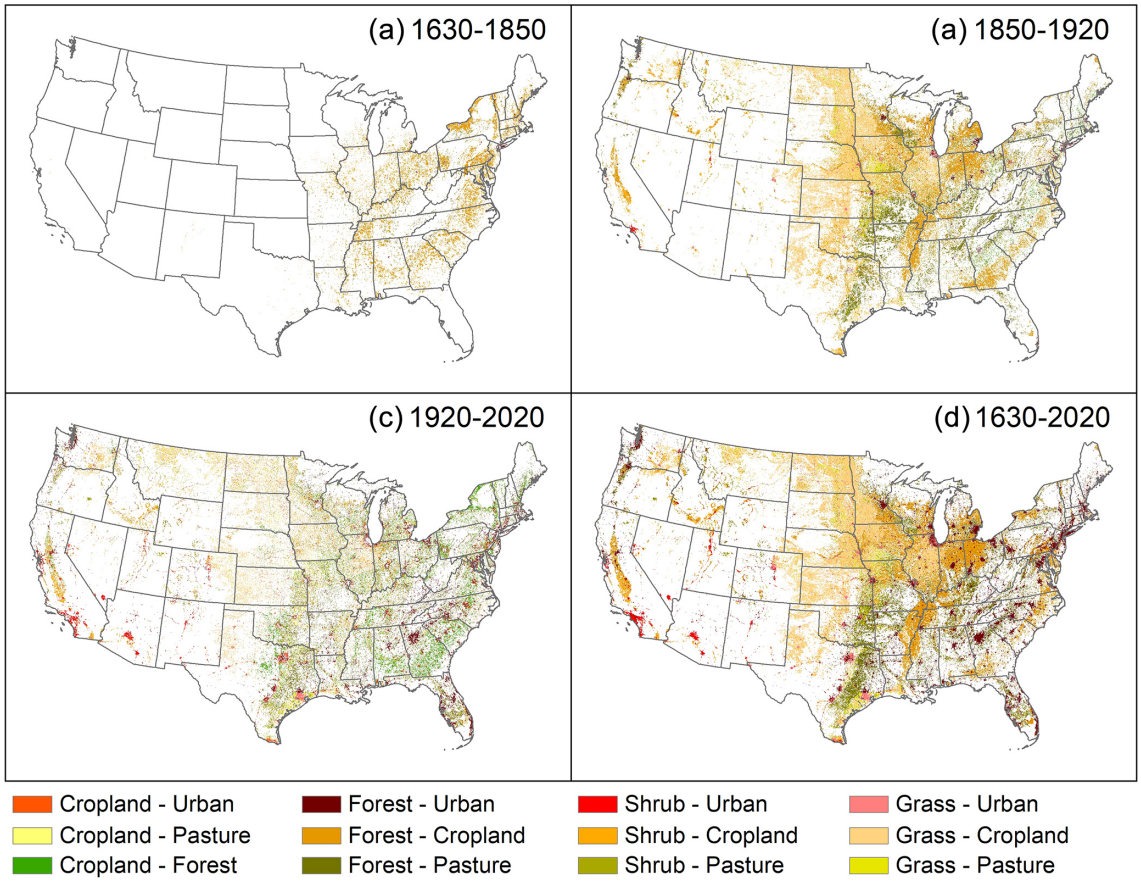

Figure 7. Land transition (1 km × 1 km spatial resolution) between 1630 and 1850 **(a)**, 1850 and 1920 **(b)**, 1920–2020 **(c)**, and 1630 and 2020 **(d)** in the conterminous United States.

and service industry resulted in agricultural labor and cropland area reduction. As the demand for biofuel and bulk grain grew in the 2000s, cropland began to extend again, and the total cropland area in 2020 was 126.94 Mha (Fig. 6). Pasture showed an increasing trend with a slowly increasing rate during 1630–1850. It expanded more than 20 times from 1850 to 1950 and reached the maximum historical area (56.94 Mha) in 1959. The total pasture area in the CONUS kept relatively stable and decreased slowly in the following 70 years (Fig. 6).

Forest was the dominant LULC type in the CONUS before the colonial era, which accounted for about 47 % of the total land area. The trends in forest area were contrary to that of agricultural land before 1920. During 1630–1850, the national total forest loss was 40.83 Mha (Fig. 6). Over the second period (1850–1920), forest area decreased by 83.02 Mha because of agricultural land expansion, lumber cutting, and fuelwood consumption. In the third period (1920–2020), forest area has been relatively stable through forest management and planting (Fig. 6).

3.2 Land use and land cover transitions during 1630–2020

The changes in the LULC area only reflected its quantitative changes. However, the LULC transition map further illustrates the spatial conversion distribution between two LULC types (Fig. 7). Over the past 390 years, cropland expansion by occupying forest, shrub, and grassland was the primary LULC change characteristic (Fig. 7). The natural land loss was mainly distributed in the North Central (e.g., Ohio, Indiana) and southern states such as Tennessee, Texas, Alabama, and Georgia (Fig. 7d). Cropland reclamation encroached 54.38 Mha (15.00 % of total forest in 1630) of forest and 68.56 Mha (19.60 % of total shrub and grassland in 1630) of grassland and shrub. Meanwhile, 37.76 Mha of forest and 11.15 Mha of shrub and grassland were converted to pasture. Moreover, urban land occupied more than 33.90 Mha of forest and 11.57 Mha of grassland and shrub (Table 3). In the early period (1630–1850), forest converted to cropland was the dominant LULC transition type, which was mainly distributed in the eastern US (Fig. 7a). The US experienced the most dramatic LULC conversion with large forest and grassland loss in the North Central and Great Plains during 1850–

1920 (Fig. 7b). Cropland expansion encroached 56.21 Mha of forest and 59.01 Mha of grassland, and pasture development also occupied more than 27.61 Mha of forest (Table 3). Furthermore, urban land expansion and abandoned cropland converted to forest (22.35 Mha) distributed in the northeastern and southern states were the essential feature of LULC changes between 1920 and 2020 (Fig. 7c).

3.3 Land use and land cover changes during 1630–2020 at the regional level

Given the differences in natural environmental conditions and social–economic development, land use and land cover changes showed spatial heterogeneity in the CONUS during 1630–2020. Since 1630, the South Central experienced the most intensive urban land expansion (10.62 Mha), followed by the North Central (10.28 Mha), Southeast (7.38 Mha), and Northeast (6.00 Mha), respectively (Fig. 8a). Rapid cropland expansion first occurred in the North Central, Northeast, South Central, and Southeast in the 1830s. Cropland in the Intermountain region and the Great Plains began to develop after 1860. The trends of cropland in eight regions except South Central and Southeast were consistent with the national total. Over the past four centuries, the North Central region had the largest cropland expansion area (46.01 Mha), followed by the Great Plains (31.41 Mha) and the South Central (20.10 Mha) (Fig. 8b). Cropland in the South Central and Southeast had decreased by 4.91 and 12.44 Mha since the 1930s due to the increasing urbanization pressures and low cropland profitability.

Similar to cropland, the Northeast was the first to develop pasture. The pasture experienced a rapid expansion during 1790–1950, reached the maximum historical area (4.56 Mha) in the 1950s, and then gradually decreased (Fig. 8c). After the 1900s, the South Central had the largest pasture area. The maximum historical area was 21.07 Mha in 1950 and accounted for 37 % of the national total. However, the pasture area in the North Central began to decrease in 1960 and 11.17 Mha of pasture was left in 2020 (Fig. 8c).

Agricultural land encroachment, land clearing, and deforestation resulted in forest loss in eight regions (Oswalt et al., 2014, 2019). In the past four centuries, North Central lost the most forest area (36.12 Mha), followed by South Central (24.85 Mha). During 1850–1920, the forest area decreased rapidly in the North Central (24.96 Mha), South Central (29.39 Mha), Southeast (14.01 Mha), and Northeast regions (6.50 Mha). Most of the lost forest was converted to cropland and pasture (Fig. 8d). Since the 1920s, the regional forest area has been relatively stable with small fluctuations. Notably, the forest land recovered gradually, especially in the Northeast, South Central, and Southeast. Compared with the 1920s, the total forest area in the Northeast increased by 6.87 Mha (Fig. 8d).

3.4 Comparison with other datasets

3.4.1 State-level land use and land cover area comparison

We compared the state-level urban, cropland, pasture, and forest areas using data derived from ERS, HISDAC, HYDE, NLCD, LUH2, and YLmap with the newly developed LULC dataset. Generally, our data match well with the data used for comparison (Fig. 9). The urban land area from this study is higher than the ERS data (Fig. 9a; $R^2 = 0.93$, slope $= 0.61$) because ERS urban land only includes the densely populated areas with at least 50 000 people (urbanized areas) and densely populated areas with 2500 to 50 000 people (urban clusters). In contrast, HISDAC built-up areas data are higher than our data (Fig. 9a; $R^2 = 0.88$, slope $= 1.34$), especially in Georgia, New York, North Carolina, Ohio, and Tennessee. This is because the HISDAC data are rebuilt using detailed property records and have a relatively coarse resolution (Leyk et al., 2020). The cropland area derived from this study is consistent with NLCD (Fig. 9b; $R^2 = 0.99$, slope $= 1.02$) and YLmap (Fig. 9b; $R^2 = 0.99$, slope $= 0.93$). Nevertheless, the ERS cropland is higher than our data (Fig. 9b; $R^2 = 0.96$; slope $= 1.26$) because the ERS cropland here includes the area of the cropland harvested area, crop failure, cultivated summer fallow, cropland used for pasture, and idle cropland. The coefficients of determination between our pasture acreages and NLCD (Fig. 9c; $R^2 = 0.93$, slope $= 1.02$) and HYDE (Fig. 9c; $R^2 = 0.87$, slope $= 0.99$) are higher than 0.87. For the forest, both NLCD and LUH2 data are lower than our data, especially in the Rocky Mountain states (Fig. 9d; $\text{slope}_{\text{NLCD}} = 0.72$, $\text{slope}_{\text{LUH2}} = 0.66$). The differences in definition and data development method could result in LULC area differences for both pasture and forest (Tables S1–S4), making it hard to compare. For example, the LUH2 forest area in the Rocky Mountain states is lower than our data and NLCD because they applied biomass density data to determine the forest extent. Though there still are some uncertainties, the comparison results show that the newly developed dataset can provide a relatively accurate LULC area at the state level.

3.4.2 Comparison with cropland census data at the county level

An accurate cropland map is quite critical for historical LULC reconstruction. We compared our data with county-level census data to assess the accuracy. This study's spatial pattern of cropland proportion (i.e., cropland area/county area) is close to the census data in 1850, 1920, 1959, and 2002 (Fig. 10). In 1850, both the newly developed cropland and census data showed high cropland density in the Black Belt, New England, and the North Central regions. In contrast, our data were higher in North Central, the east of Virginia and North Carolina, and the south of Georgia (Fig. 10). Cropland derived from this study was higher than the census

Table 3. Net land use and land cover change during 1630–1850, 1850–1920, 1920–2020, and 1630–2020. Unit: Mha.

LULC transition type		1630–1850	1850–1920	1920–2020	1630–2020
Cropland to others	Cropland to urban	0.00	0.46	10.92	0.00
	Cropland to pasture	0.00	1.58	9.09	0.00
	Cropland to forest	0.00	3.67	22.35	0.00
	Sub-total	0.00	5.71	42.36	0.00
Others to cropland	Pasture to cropland	0.00	0.27	0.97	0.00
	Forest to cropland	25.91	56.21	11.40	54.38
	Grassland to cropland	1.06	59.01	18.47	62.60
	Shrub to cropland	0.02	5.43	3.21	5.96
	Sub-total	26.99	120.92	34.05	122.94
Others to pasture	Forest to pasture	2.44	27.61	15.24	37.76
	Grassland to pasture	0.07	5.78	3.00	9.10
	Shrub to pasture	0.00	0.68	1.56	2.05
	Sub-total	2.51	34.07	19.80	48.91
Others to urban	Forest to urban	0.76	4.07	18.01	33.90
	Grassland to urban	0.03	1.60	2.81	7.56
	Shrub to urban	0.00	4.07	2.71	4.01
	Sub-total	0.79	9.74	23.53	45.47

Note: the area of LULC transition is calculated based on the intersection result of LULC data in the first and last year for each period.

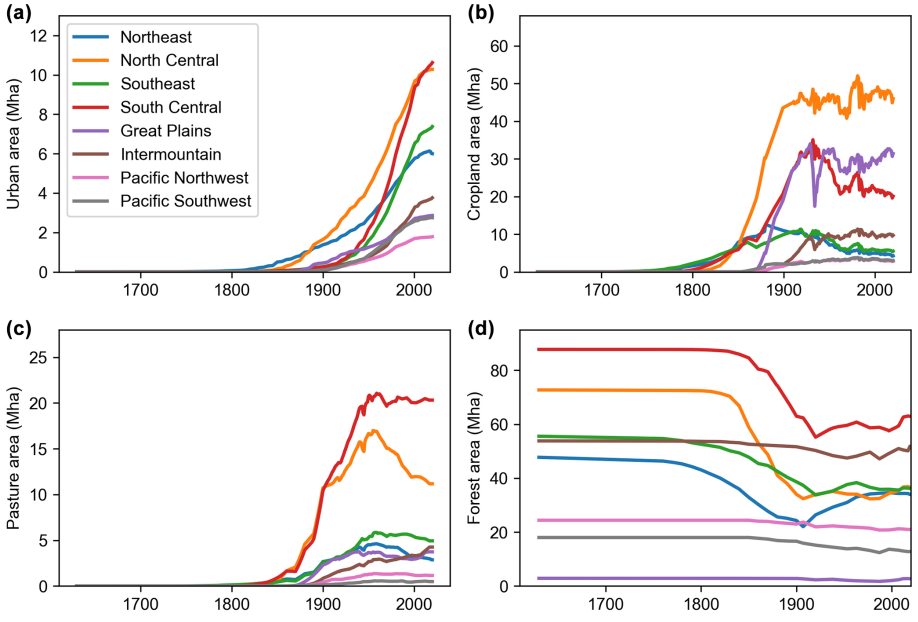

Figure 8. Changes in areas of urban land **(a)**, cropland **(b)**, pasture **(c)**, and forest **(d)** in different geographic regions during 1630–2020.

data in the Atlantic coast, the Mississippi Alluvial Plain, the northwest of Texas, the west of Oklahoma, and California in 1920, 1959, and 2002. However, the cropland proportion in the Appalachian Mountains and the south of the Great Plains was lower than the census data (Fig. 10). This underestimation may result from the low cropland fraction in satellite data, because it is difficult for satellite data to identify the small area cropland patch in the mountain region and classify the pasture or grassland with cropland in the south of the Great Plains. Moreover, both datasets showed the cropland expansion in the North Central, the Great Plains, the Mississippi Alluvial Plain, and California between 1850 and 2002. The cropland abandonment can also be found in the Appalachian Mountains between 1920 and 2002. The statistical comparison also shows that our data fit well with the census data in 1920 ($R^2 = 0.68$), 1959 ($R^2 = 0.89$), and 2002

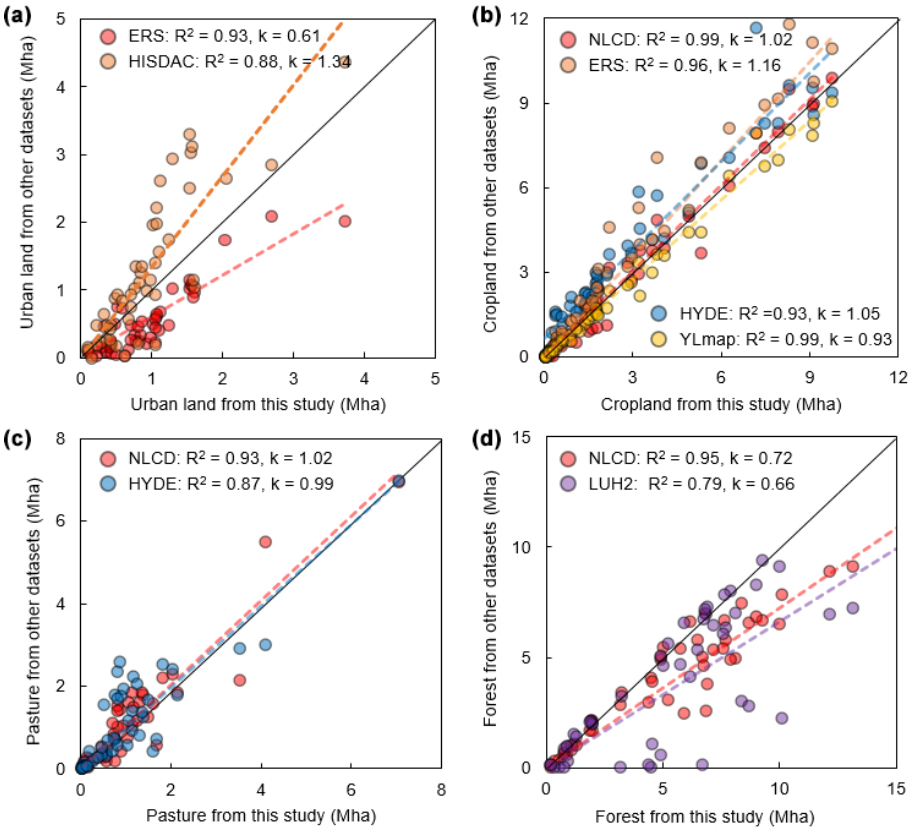

Figure 9. Comparison of the average urban **(a)**, cropland **(b)**, pasture **(c)**, and forest **(d)** area in each state among National Land Cover Database (NLCD), Historical Settlement Data Compilation (HISDAC), Economic Research Service (ERS), Yu and Lu (2017) cropland (YLmap), History Database of the Global Environment (HYDE), Land Use Harmonization (LUH2), and this study in the overlapping years during 2000–2020. This study: 2000–2020; NLCD: 2001, 2003, 2006, 2008, 2011, 2013, 2016, and 2019; HISDAC: 2000, 2005, 2010, and 2015; ERS: 2002, 2007, and 2012; HYDE: 2000–2017; YLmap: 2000–2016; and LUH2: 2000–2019.

($R^2 = 0.91$) (Fig. A6). Overall, the newly developed cropland has a relatively accurate spatial pattern and proportion.

3.4.3 Comparison with NLCD at the grid level

The spatial patterns of urban, cropland, pasture, and forest in this study are close to the satellite-based data from NLCD, and most grid cells have a relatively small difference between 2001 and 2019 (Fig. 11). Our results have a higher urban land fraction in the NLCD low urban density area, but the difference in 87 % of urban grids is smaller than 10 %. Cropland with a positive difference is mainly distributed in the Northeast, Alabama, and Missouri, in which 65.95 % of grids have slight differences with less than 10 % (Fig. 11); 37.19 %of grids have negative difference values and are mainly located in states with high cropland proportions. Moreover, most states in our data have a lower pasture fraction than NLCD data, except in Oklahoma, Arkansas, Texas, and Georgia, and the grid cells with negative differences account for 39.82 %. The reconstructed forest shows a higher density than NLCD in the South, Pacific Coast, and Great Lakes. It underestimates the forest fraction in the central states, such as Missouri, Kentucky, and Ohio. There are 58.80 % grids whose differences are relatively small and with a range from −10 % to 20 % (Fig. 11).

4 Discussion

4.1 Comparison with the previous datasets

Compared with the ERS and HYDE data, the reconstructed urban land area was higher (Fig. 12a), attributed to the definition differences with NLCD. The ERS urban area includes the densely populated areas with at least 50 000 people (urbanized areas) and densely populated areas with 2500 to 50 000 people (urban clusters). The total urban land area from HISDAC data was higher than the newly developed data in the recent four decades (Fig. 12a). Because the HISDAC built-up area dataset was developed by using the detailed property records data at a relatively coarse resolution (Leyk et al., 2020), some small-scale built-up land cannot be identified using satellite images and NLCD may underestimate the total urban land area. Moreover, the HISDAC built-

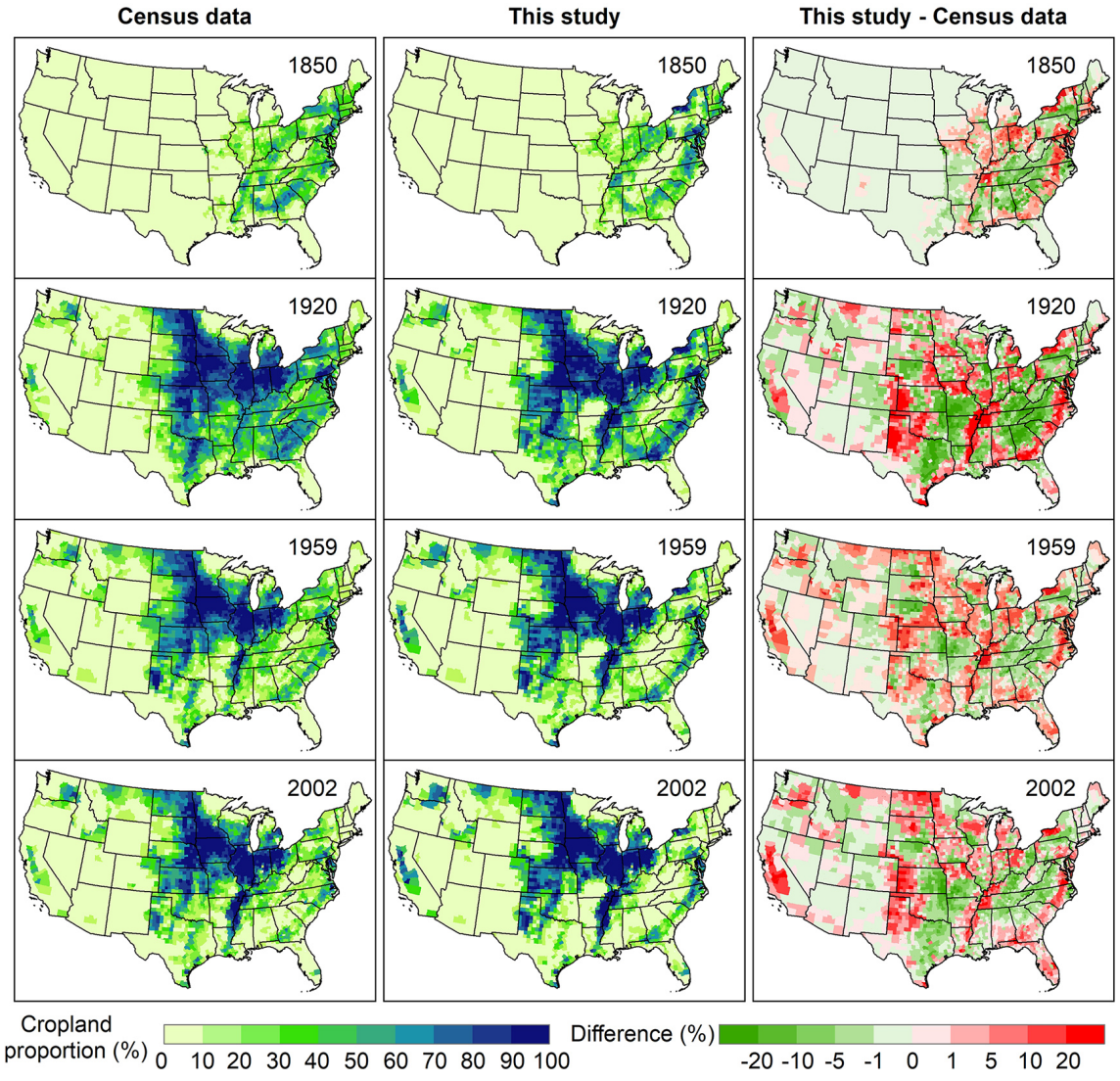

Figure 10. Spatial comparison of county-level cropland proportion between our reconstruction and census data in 1850, 1920, 1959, and 2002. First column: cropland proportion from census data, second column: cropland proportion derived from this study, and third column: cropland proportion difference between this study and census data.

up areas underestimated the total urban area in the early years due to the high missing rate of property records (Leyk et al., 2020). Therefore, our data may also underestimate the total urban land area because we applied the trend of HISDAC between 1810 and 2001. The spatial pattern of Boolean type urban land was consistent with the Sohl et al. (2016) data and was mainly distributed in the area near the city, road, and railway (Fig. 13). The spatial allocation rule determined that the grid with a high probability of occurrence would be allocated first, which may underestimate the developed land in the rural area (Verburg and Overmars, 2009; Yang et al., 2020). Though some uncertainties in the urban data exist, we provided a long-term description of urban land with higher resolution and consistency for the CONUS.

The reconstructed cropland area was close to NLCD in the 2000s (Fig. 12b). Our data and YLmap applied the cropland harvested area to estimate the historical cropland area and showed the same trend during 1850–2016 (Fig. 12b). The cropland area derived from ZCmap and ERS was higher than our data over the research period (Fig. 12b) because cropland harvested, crop failure, cultivated summer fallow, cropland use for pasture, and idle cropland all counted (Zumkehr and Campbell, 2013; Bigelow and Borchers, 2017). The area trend between 1630 and 1879 was close to HYDE because we used its cropland per capita trend (Fig. 12b). Spatially, four fractional cropland maps show the similar state and expansion patterns. The highest cropland density can be found in the Corn Belt, Central California, and Mississippi Alluvial Plain in the 2000s. Meanwhile, cropland expansion ini-

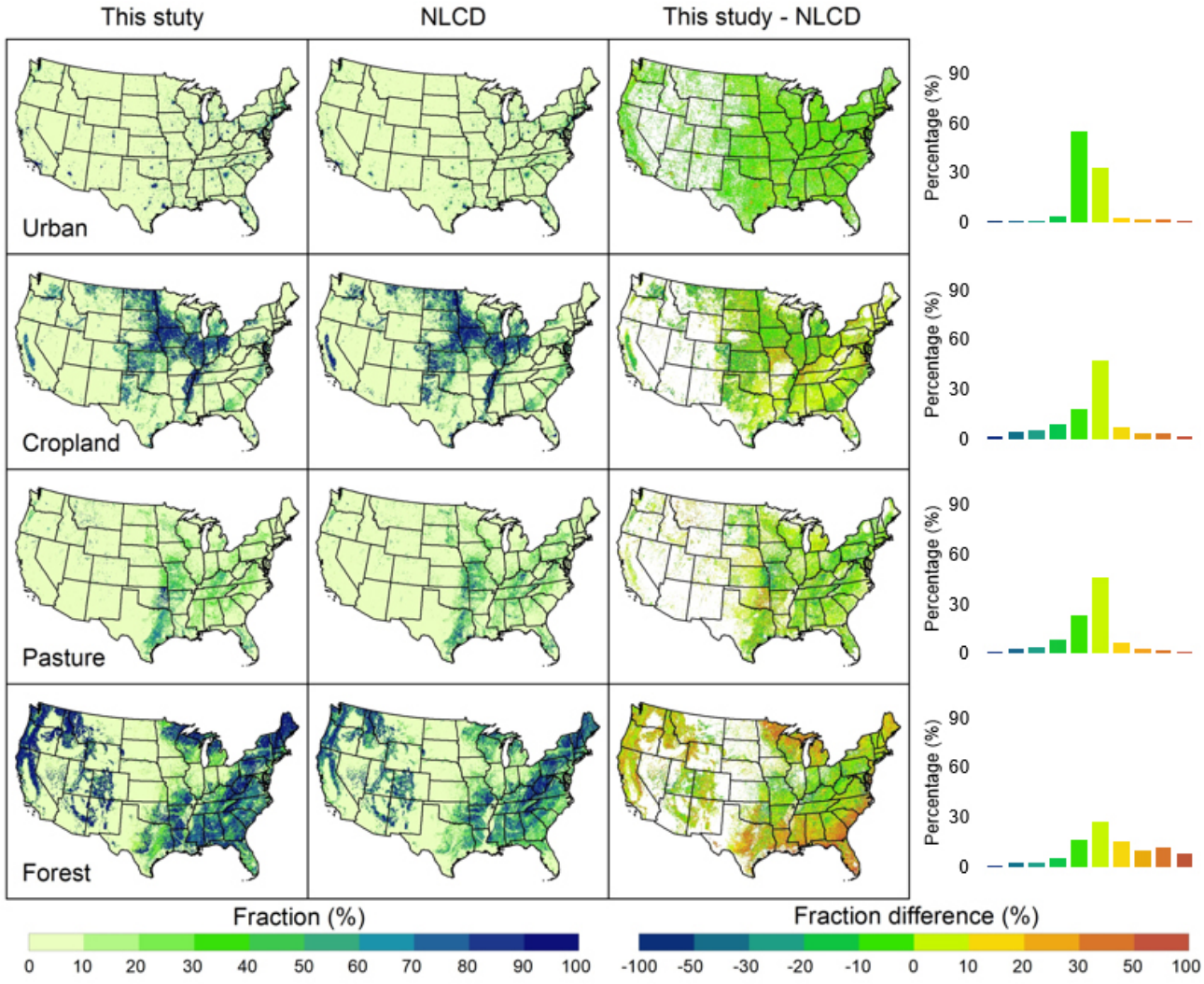

Figure 11. Spatial comparison between our reconstruction and satellite-based urban, cropland, pasture, and forest. First column: reconstructed data in this study (average between fraction of 2001, 2003, 2006, 2008, 2011, 2013, 2016, and 2019), second column: satellite-based data (average fraction of 2001, 2003, 2006, 2008, 2011, 2013, 2016, and 2019), third column: difference between first column and second column, and fourth column: distributions of fraction difference between our reconstructed database and satellite-based data.

tially occurred in the east of Mississippi River, then moved to the Midwest and the Great Plains between 1850 and 1920 (Fig. 14). Our results can reflect the cropland abandonment in New England, the South, and the Southeast since the 1920s, which is consistent with previous studies (Reuss et al., 1948; U.S. Department of Agriculture and Economic Research Service, 1974; Foster, 1992). Moreover, the newly developed cropland improved the spatial resolution compared with HYDE and ZCmap, making it possible to catch more detailed information (Fig. 15). In YLmap, there are some coarse grids in the early years (Fig. 15) because they applied HYDE data to reconstruct the cropland expansion and abandonment (Yu and Lu, 2018). Our data were processed at 1 km resolution and fixed this problem (Fig. 15). Compared with the above cropland data, our product has higher spatial resolution and more extended temporal coverage, making it capable of depicting the cropland dynamics better in the CONUS over the past four centuries.

To the best of our knowledge, accurate temporal and spatially explicit data are still lacking to describe the pasture dynamics for the CONUS. This study set the state-level pasture area from the National Resource Inventory (NRI) as the baseline data for historical pasture reconstruction, which made our data more reliable than HYDE. During 2001–2020, the total national area of pasture located in non-federal land ranged from 48 to 53 Mha, which was close to the NLCD (53 Mha) and HYDE (52 Mha) (Fig. 12c). We also found that NLCD pasture/hay decreased during 2001–2016, while NRI pasture remained relatively stable. The likely reasons for NLCD pasture/hay loss include normal crop cycling and

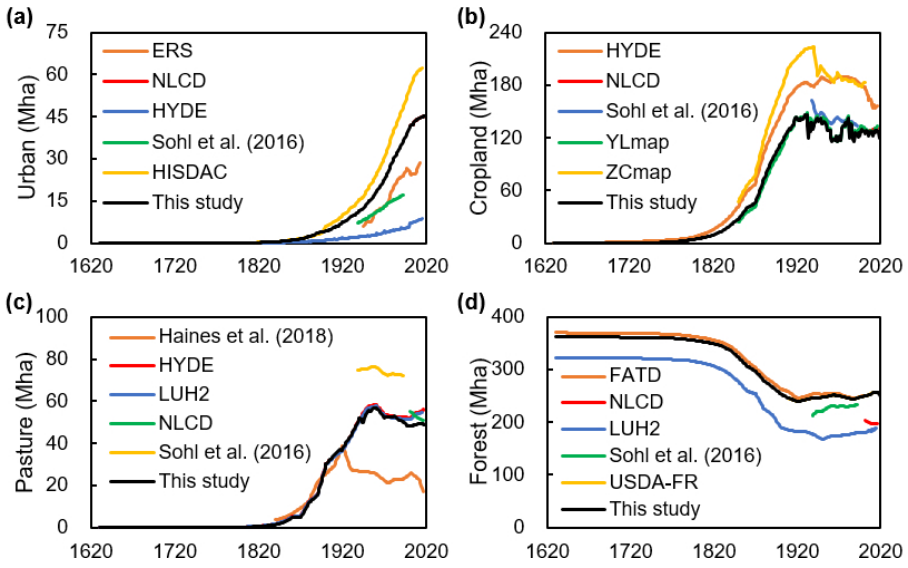

Figure 12. Comparison with other datasets for the conterminous United States: urban land **(a)**, cropland **(b)**, pasture **(c)**, and forest **(d)**. NLCD: National Land Cover Database, HYDE: History Database of the Global Environment, HISDAC: Historical Settlement Data Compilation, ERS: Economic Research Service, YLmap: Yu and Lu (2017) cropland density, ZCmap: Zumkehr and Campbell (2013) historical fractional cropland areas, LUH2: Land Use Harmonization, FATD: Forest Area Trend Data, and USDA-FR: USDA Forest Resources of the United States of 2017.

more permanent conversion (Homer et al., 2020). The difference in pasture trends between NRI and NLCD may result from the definitional difference (Table S6). Nevertheless, for Haines et al. (2018), pasture only includes hay, making it significantly lower than our result (Fig. 12c). The ERS data also provided grazing land area, but the rangeland and pasture were not separated (Bigelow and Borchers, 2017). The application of the HYDE pasture per capita trend made our result close to it and reached the maximum historical value in the 1950s (Fig. 12c). The three maps all show the highest pasture density in eastern Texas, Oklahoma, and Missouri (Fig. 16). At the regional scale, the spatial patterns of pasture from this study are close to the HYDE and LUH2 data, but our data can characterize the historical changes of pasture with higher spatial resolution than current LULC products (Fig. 17).

We used the inventory-based datasets (FATD and USDA-FR) to reconstruct the historical forest area. Compared with satellite-based forest (NLCD), Sohl et al. (2016), and LUH2 data, the total forest area in our data is higher. This area difference mainly resulted from the differences in forest definition. For example, NLCD and Sohl et al. (2016) define forest as areas dominated by trees generally greater than 5 m tall and greater than 20 % of total vegetation cover, higher than that in our forest definition (forest cover greater than 10 %) (Sohl et al., 2016; Homer et al., 2020; Table S7). Moreover, the forest in LUH2 is determined by the vegetation biomass density and country-level forest area (Hurtt et al., 2020), underestimating the forest distribution in the area with low biomass density. Spatially, our data and LUH2 can describe the high density in the eastern US and Pacific Coast area, but LUH2 underestimates the forest fraction in Rock Mountain and Texas (Figs. 18 and 19). Our data fixed the above problem and improved the spatial resolution from 0.25° to 1 km. Meanwhile, the newly developed forest data have good performance in capturing forest dynamics. For example, previous studies reported deforestation in southern Michigan and forest cutting for agriculture and fuel in Virginia during the early settlement period (Garrison, 2012; Mergener et al., 2014), also shown in our maps during the 19th century (Fig. A5). Forest loss during the westward expansion period can be captured in the Northeast, Midwest, and Great Plains (Figs. 18, A5). The LULC conversion map can reveal the forest regrowth on much cutover and abandoned agricultural land in the Northeast and Southeast since the 20th century (Foster et al., 1998; MacCleery, 2011) (Figs. 10, A5).

4.2 Drivers of land use and land cover changes

Agricultural land expansion and natural vegetation loss (forest, grassland, and shrub) area is the primary characteristic of LULC change in the CONUS over the past four centuries. The complex interactions among land suitability, climate, population, transportation, agricultural technologies, and policy shaped the contemporary LULC pattern. In the colonial era, the migration of Europeans into the Northeast and mid-Atlantic converted the eastern forests to cropland and pasture (Waisanen and Bliss, 2002). More than 90 % of people lived in the east of the Appalachian Mountains, and most farms were subsistence in this period. The forced migration of slaves contributed to the plantation agriculture

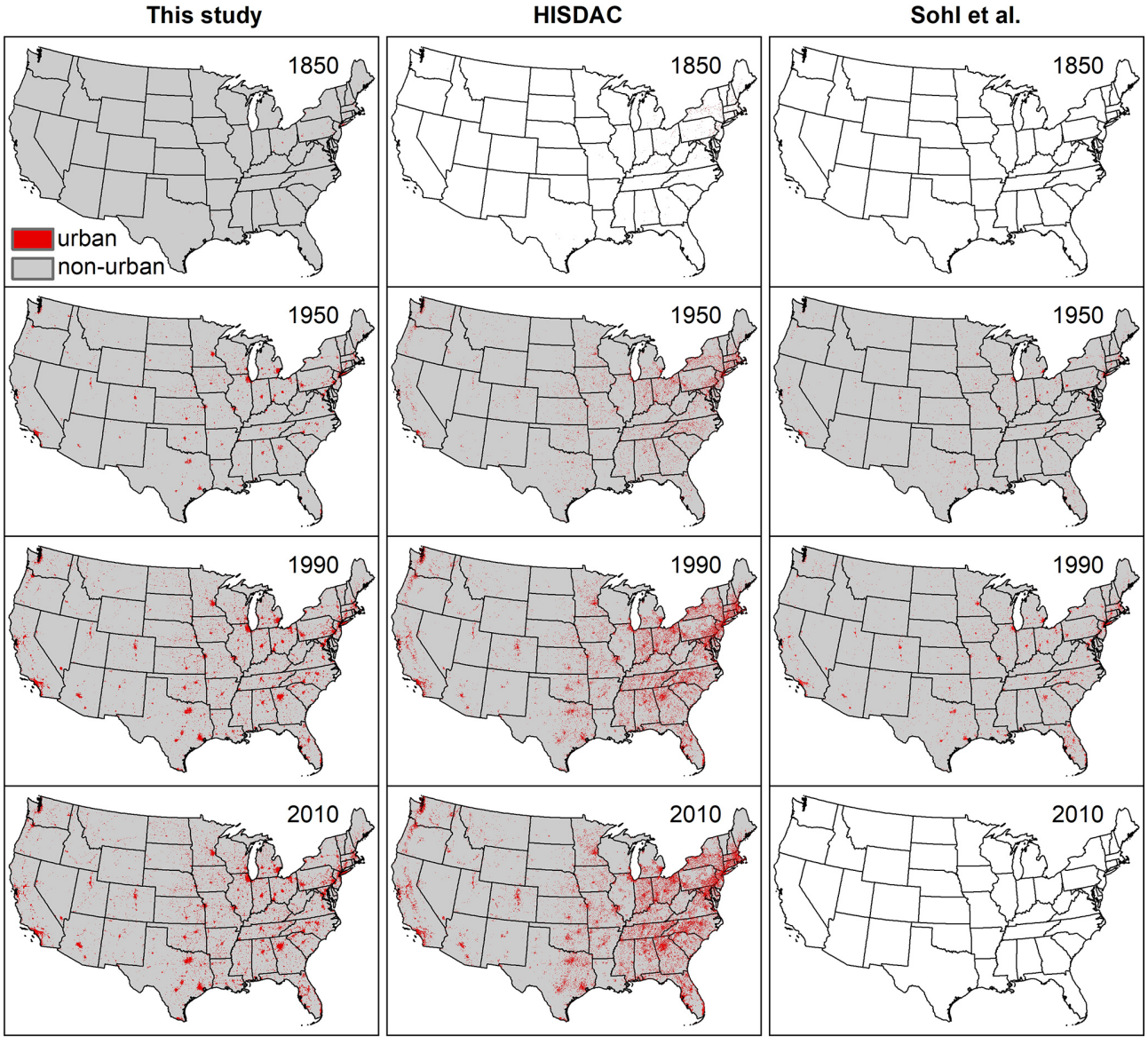

Figure 13. Comparison of urban land maps among three datasets for the conterminous United States: this study (left column), Historical Settlement Data Compilation (HISDAC) map (central column), and Sohl et al. (2016) map (right column).

expansion in Virginia, Maryland, South Carolina, and the Black Belt. After the new nation was established, numerous lands like Louisiana, Florida, Texas, Oregon, and New Mexico were acquired during 1800–1860 (Fretwell et al., 1996). The westward movement opened new areas for agricultural development. With the building of canals and inland waterways, agricultural products from the cropland developed west of the Appalachians could be brought to the market (Meinig, 1993). In the second half of the 19th century, the rapid population growth and food demand resulted in cropland expansion because farmers needed to reclaim another 3–4 acres to feed one person (MacCleery, 2011). After the 1920s, cropland, pasture, and forest area became relatively constant despite the growing population. The applications of hybrid crops and fertilizers and the increasing number of motor vehicles and farmer tractors improved agricultural productivity, which played an essential role in stabilizing cropland area (Waisanen and Bliss, 2002; MacCleery, 2011). Cropland abandonment in the East was affected by the fluctuations in crop prices, changes in labor markets and competition from the high productivity in the Midwest (Hart, 1968; Williams, 1992; Bigelow and Borchers, 2017). The reversion of marginal cropland in the East and large-scale tree planting in the South contributed to the forest recovery (Clawson, 1979; Smith et al., 2001; Thompson et al., 2013). For example, many croplands in the South were abandoned following

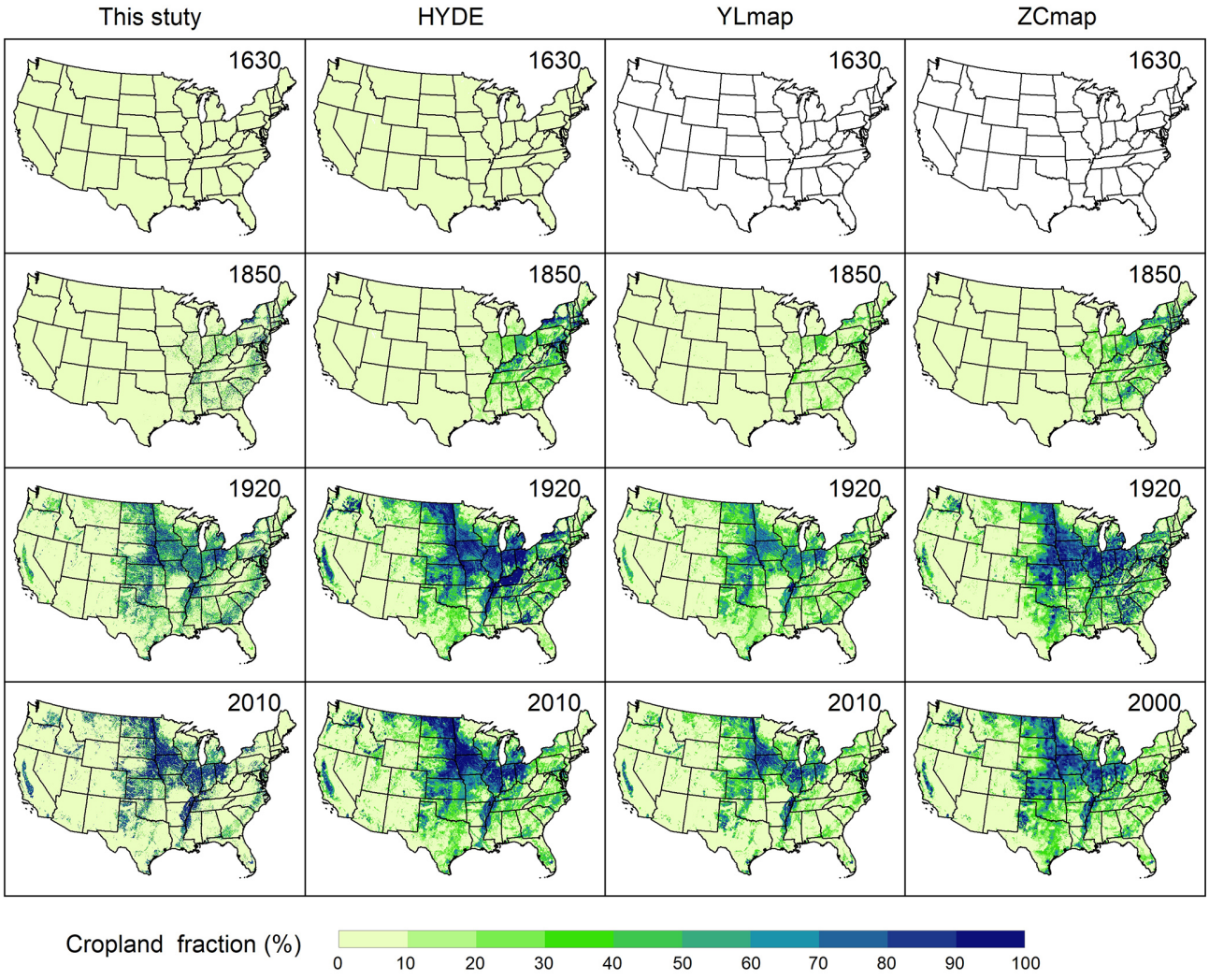

Figure 14. Comparison of cropland maps among four datasets for the conterminous United States: this study (first column), the History Database of Global Environment (HYDE), Yu and Lu (2017) cropland density (YLmap), and Zumkehr and Campbell (2013) historical fractional cropland areas (ZCmap).

the disintegration of the post-bellum sharecropping system and later converted to plantation forests (Hart, 1968), and the plantation forest area increased from near zeros in the 1930s to 27 Mha in 2017 (Chen et al., 2017; Oswalt et al., 2019). Climate change also impacts the LULC change. For example, the Dust Bowl in the 1930s led to widespread crop failure in the Great Plains (Heimlich and Daugherty, 1991). Land marked by crop failure due to severe drought, extensive flooding, or wet weather has ranged between 5 and 22 million acres since 1949 (Bigelow and Borchers, 2017).

4.3 Uncertainties and future perspectives

This study provides a 4-century LULC dataset at an annual time step and 1 km × 1 km spatial resolution for the CONUS. However, some uncertainties may affect the accuracy of this dataset. For instance, both the reliability of input data and the harmonization method are critical for the historical LULC area reconstruction. Most census data used in this study were recorded at 4- to 10-year intervals, making some interannual fluctuations impossible to capture. The rebuilt state-level LULC area is also coarse if there are significant spatial shifts (e.g., cropland abandonment in some counties but reclamation in others) for an LULC type. Moreover, the definitional differences among datasets increased the difficulties and uncertainties in the harmonization process. Though we tried to gather the most reliable LULC datasets, the definitions of LULC vary (Tables S5–S7). The definitions of four LULC types do not belong to a universal classification system, making it hard to process the total area, and a post-processing step needs to be conducted. For the urban land, HISDAC built-up area was higher than that from the NLCD dataset. However, the urban land area change rates were close

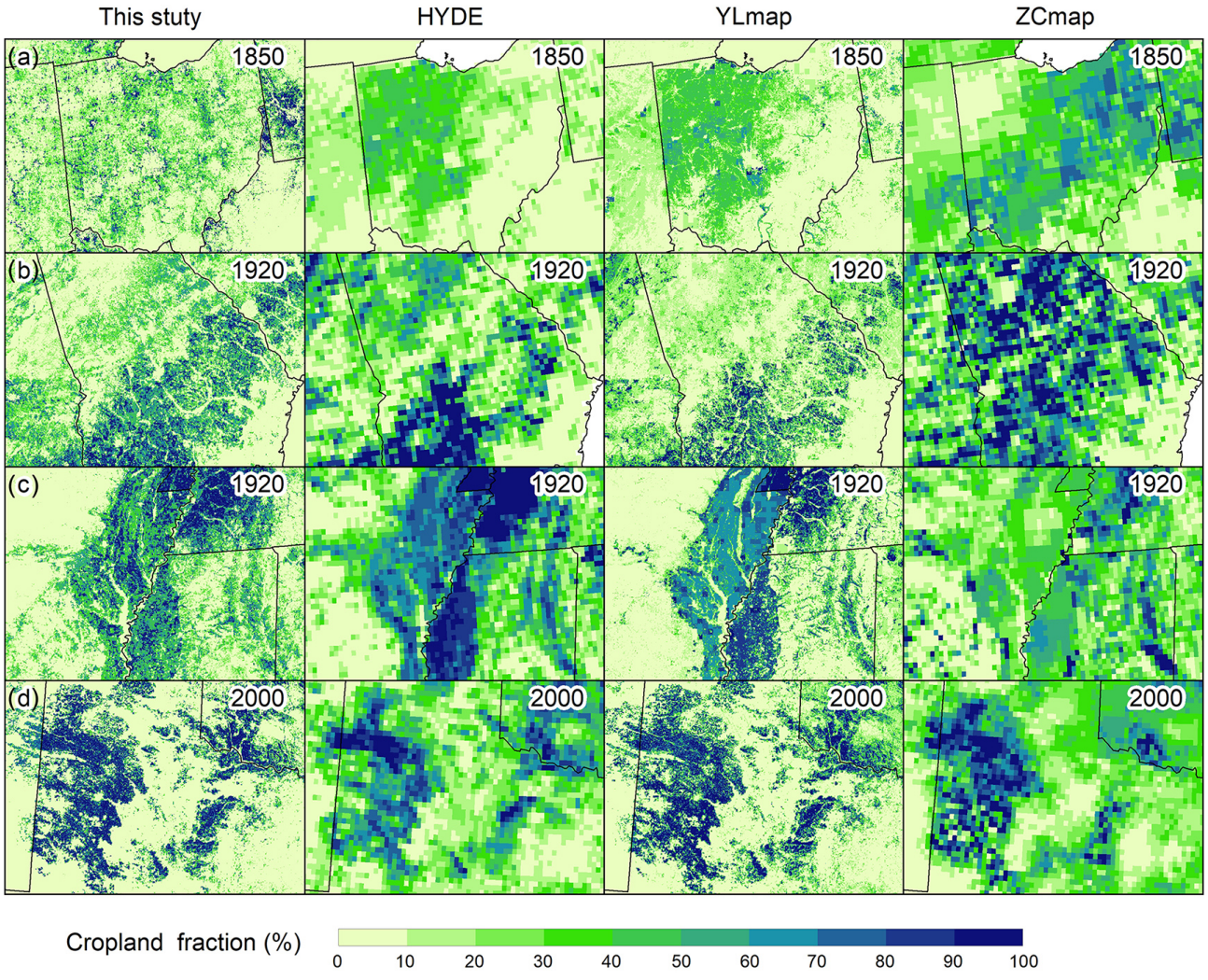

Figure 15. Visual comparison between our cropland data and the History Database of Global Environment (HYDE), Yu and Lu (2017) cropland density (YLmap), and Zumkehr and Campbell (2013) historical fractional cropland areas (ZCmap) in four different sites **(a–d)**. The locations of image center points are as follows: **(a)** Ohio (83.05° W, 40.17° N), **(b)** Georgia (83.58° W, 32.77° N), **(c)** Arkansas (90.56° W, 34.76° N), and **(d)** Texas (100.92° W, 32.81° N).

during 2001–2015, indicating that there would be minor uncertainties in combining the HISDAC data and the newly developed urban land (Fig. S8). We applied three datasets (i.e., ERS cropland harvested area, CAHA cropland harvested area, HYDE cropland) to generate the cropland area for the study period, but the definitions of cropland harvested area and cropland are different (Table S5). We conducted a data adjustment to subtract the double-cropped cropland area and optimize the cropland harvested area interannual variations, but it only resulted in little change in the cropland area (2.23 ± 1.81 %) (Fig. S9). The relative difference of cropland per capita change rate during 1910–2000 between HYDE3.2 cropland data and the newly developed cropland data is relatively higher than that during 2001–2017 and 1880–1909, but the values in most of the years were lower than 5 % (Figs. S10 and S11). For the pasture, the state-level mean relative difference in pasture per capita change rate between HYDE3.2 pasture data and our data is 4.89 ± 1.94 % during 1982–2017 (Fig. S13). Thus, though some uncertainties are introduced, it is acceptable to generate the historical LULC data.

More efforts are needed to generate accurate historical LULC maps for understanding the history of regional LULC changes. An accurate LULC probability or suitability surface is the key to generate spatial data. In this study, we assumed that the ANN-based LULC probability was unchanged following previous LULC simulation models (Verburg et al., 2006; Sohl et al., 2014; Liu et al., 2017). However, we found that the contemporary probability surfaces could not represent the historical LULC pattern, especially for agricultural

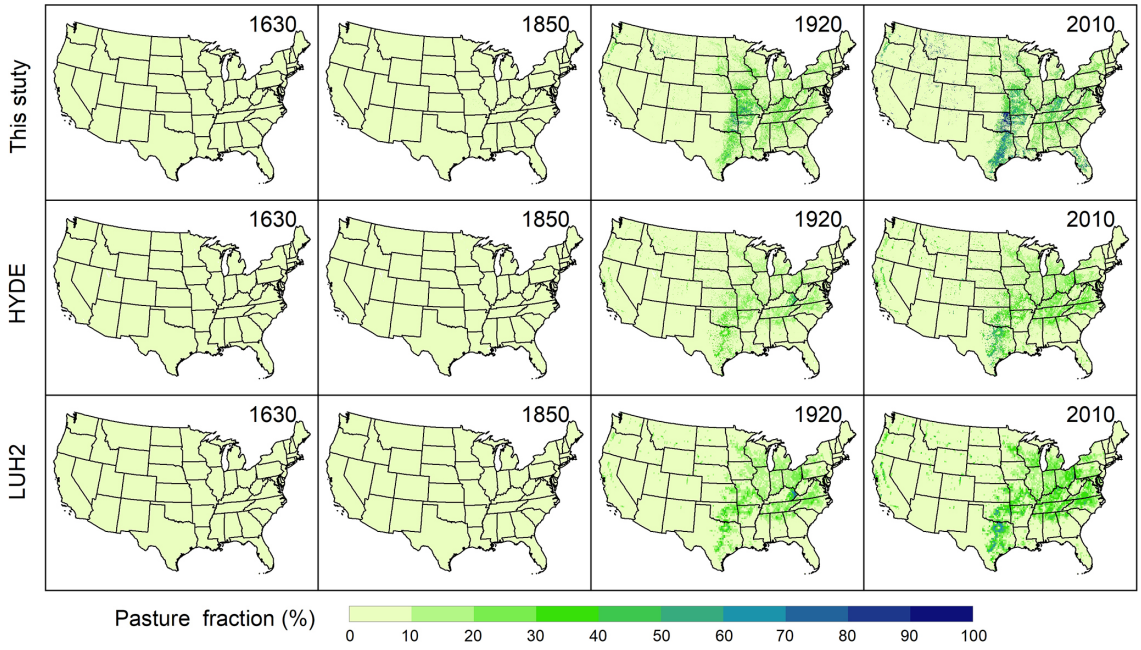

Figure 16. Comparison of pasture patterns and changes among three data products for the conterminous United States: this study (upper row), the History Database of Global Environment (HYDE) (middle row), and Land Use Harmonization (LUH2) (lower row).

land (Sohl et al., 2016). To solve the problem, we modified the LULC probability by using population density, human settlement extent, and satellite-observed LULC fraction, making it match the historical LULC pattern. The LULC pattern is highly related to that in the previous year, and the grid value is also affected by the fraction and type in the neighbor grid cells. In our spatial allocation strategy, we generate the LULC map for each year based on the LULC probability or LULC fractional data, which ignores the LULC pattern interactions between the adjacent years. Some studies generated a LULC map by allocating the LULC net change area to a base map (West et al., 2014; Liu et al., 2017; Cao et al., 2021), but such an algorithm would underestimate the LULC gross change (Winkler et al., 2021). Therefore, an improved spatial allocation strategy should be developed to simulate LULC conversion better.

The newly developed LULC dataset reconstructed the LULC history with more LULC types than ZCmap and YLmap and has higher spatial resolution than HYDE and LUH2. Our LULC data emphasize the accuracy of area change resulting from LULC conversion rather than the changes in LULC structure or attributes. For example, forest management (e.g., wood harvest and thinning) results in forest cover decreases and ecosystem function change, but the LULC type is unchanged. HYDE and LUH2 not only have a more extended cover period, but also provide more sub-types and LULC attributes. HYDE classified cropland into rain-fed rice, irrigated rice, rain-fed other crops, and irrigated other crops (Klein Goldewijk et al., 2017). LUH2 divides cropland into C_3 crops and C_4 crops and includes the wood harvest (traditional fuelwood, commercial biofuels, and industrial roundwood) and primary/secondary forest age (Hurtt et al., 2020). In the future, the LULC sub-types (e.g., tree species, crop types) and attributes (e.g., forest age, management intensity) gained through collecting them from agricultural census data and forest inventory data can be incorporated into our dataset (Thompson et al., 2013; Chen et al., 2017; Crossley et al., 2021).

5 Data availability

The land use and land cover datasets for the conterminous United States are available at https://doi.org/10.5281/zenodo.7055086 (Li et al., 2022). The annual gridded datasets (1 km × 1 km spatial resolution) with GeoTiff format include fractional and Boolean types. An Excel table is used to organize the annual urban, cropland, pasture, and forest area at the state level. A detailed data description is also provided.

6 Conclusions

This study developed spatially-explicit LULC data at a spatial resolution of 1 km × 1 km and an annual timescale in the CONUS during 1630–2020 by integrating multisource datasets. The results showed that extensive cropland and pasture expansion and natural vegetation loss occurred from 1630 to 2020 in the CONUS. New reclaimed cropland was primarily converted from forest, shrub, and grassland. Tree planting and forest regeneration increased the forest cover

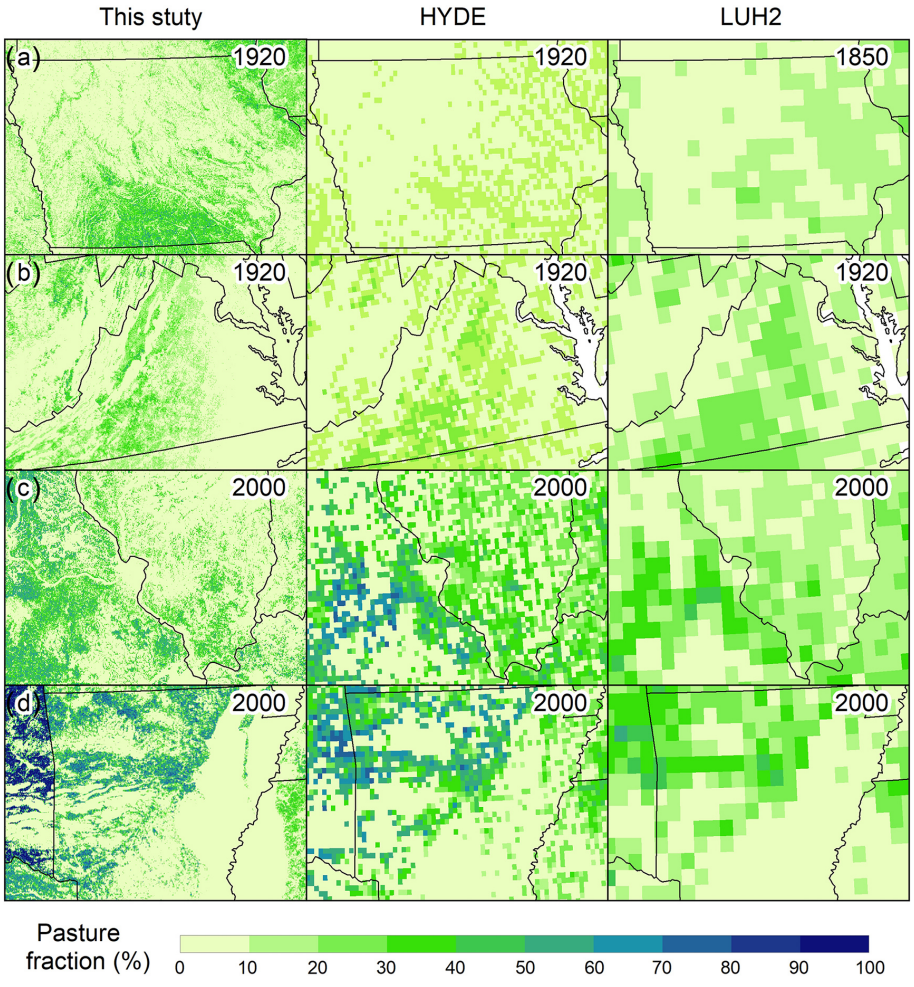

Figure 17. Visual comparison of our pasture data with History Database of Global Environment (HYDE) and Land Use Harmonization (LUH2) in four different sites **(a–d)**. The locations of image center points are as follows: **(a)** Iowa (93.64° W, 42.03° N), **(b)** Virginia (78.72° W, 37.96° N), **(c)** Illinois (90.07° W, 38.68° N), and **(d)** Arkansas (92.56° W, 34.97° N).

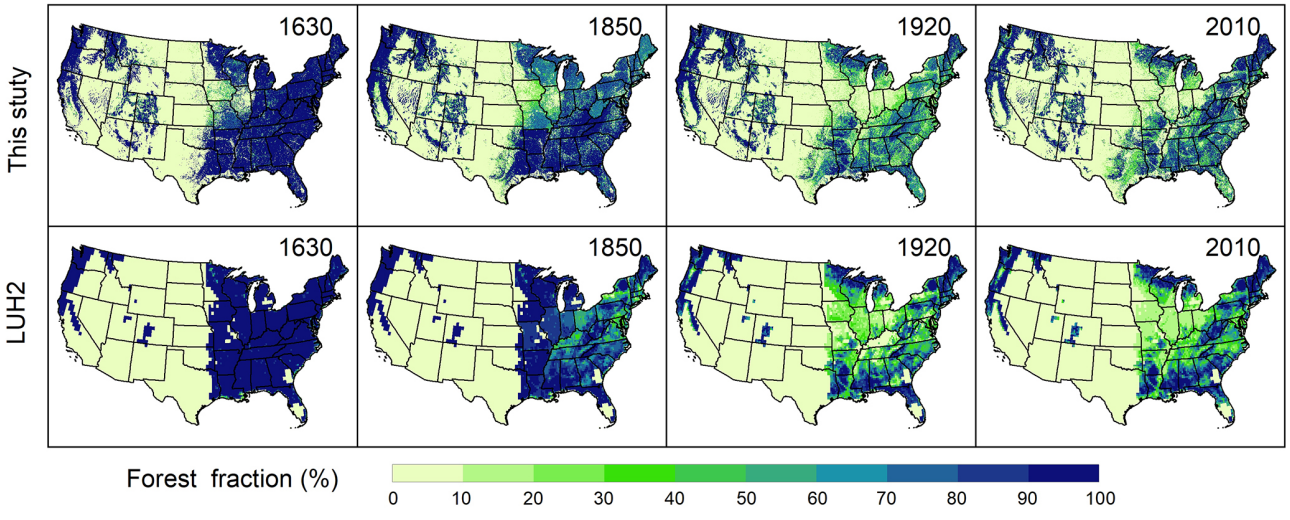

Figure 18. Comparison of forest distribution between this study (upper row) and Land Use Harmonization (LUH2) (lower row) for the conterminous United States.

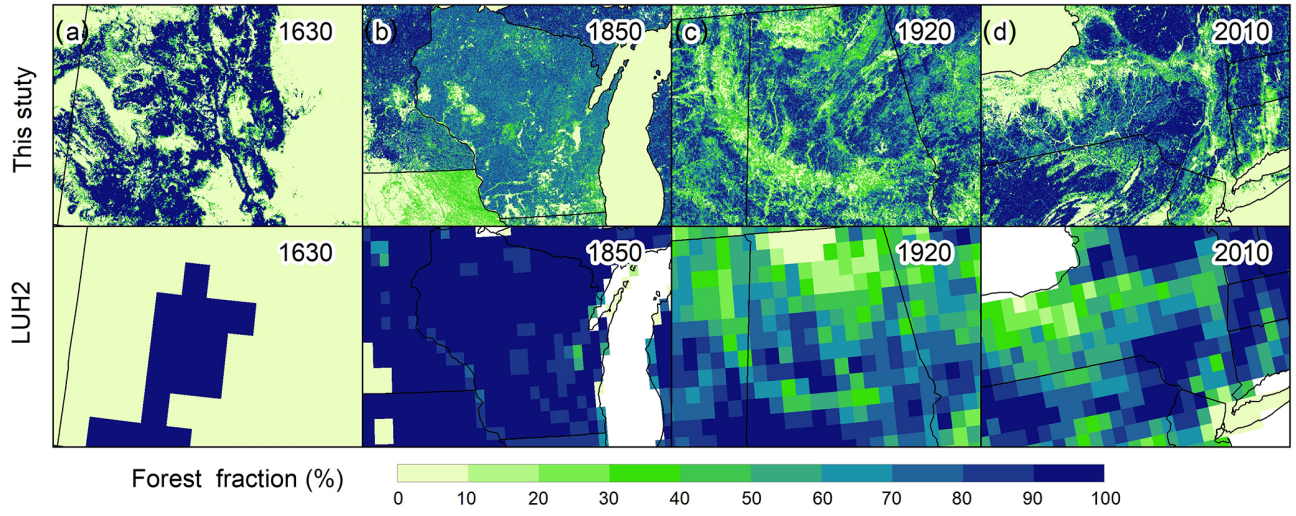

Figure 19. Visual comparison between our forest data and Land Use Harmonization (LUH2) in four different sites **(a–d)**. The locations of image center points are as follows: **(a)** Colorado (106.47° W, 38.97° N), **(b)** Wisconsin (89.85° W, 44.54° N), **(c)** Alabama (86.72° W, 33.33° N), and **(d)** New York (75.14° W, 42.21° N).

in the Northeast and the South in the recent century. Compared to other LULC datasets, our data provided more accurate information with higher spatial and temporal resolution and better captured the characteristics of LULC changes. The LULC data can be used for regional studies on various topics, including LULC impacts on regional climate, ecosystems, biodiversity, water resources, carbon and nitrogen cycles, and greenhouse gas emissions.

Appendix A

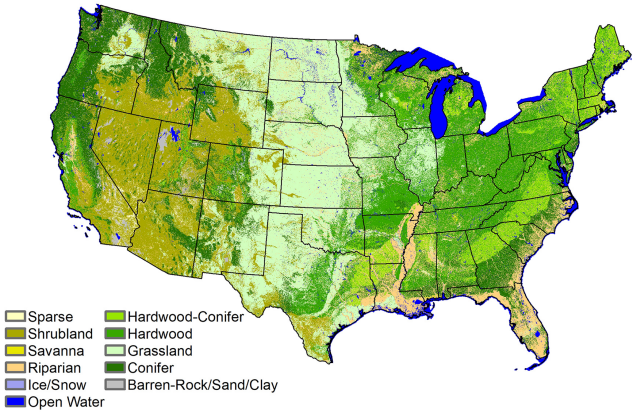

Figure A1. Vegetation type of pre-Euro-American settlement (Rollins, 2009).

Table A1. Spatially explicit variables adopted for artificial neural network (ANN) modeling.

Variable	Description	Source	Resolution
Elevation	Digital Elevation Model (DEM)	Shuttle Radar Topography Mission (STRM) (https://cgiarcsi.community/data/srtm-90m-digital-elevation-database-v4-1/, last access: 11 August 2022)	90 m
Slope	Slope calculated from DEM		
Pop	Population density	Fang and Jawitz (2018) (http://doi.org/10.6084/m9.figshare.c.3890191)	1 km
$City_{dis}$	Distance to city	https://www.sciencebase.gov/catalog/item/537d23fee4b00e1e1a484c82?community=Data+Basin (last access: 11 August 2022)	vector
$Road_{dis}$	Distance to road		vector
$Railway_{dis}$	Distance to railway		vector
$River_{dis}$	Distance to river	North America river and lakes (https://www.sciencebase.gov/catalog/item/4fb55df0e4b04cb937751e02, last access: 11 August 2022)	vector
Soil clay	Soil texture clay fraction	Soil grids 250 m v2.0, https://soilgrids.org/ (last access: 4 July 2022)	250 m
Soil sand	Soil texture sand fraction		250 m
Soil SOC	Soil organic carbon		250 m
Crop PI	Crop productivity index	Soil Survey Geographic (SSURGO) data	250 m
PPT	Annual precipitation	PRISM 30-year normal climate data (https://prism.oregonstate.edu/recent/, last access: 11 August 2022)	800 m
TMP	Mean temperature		800 m
Max TMP	July temperature		800 m
Min TMP	January temperature		800 m

Note: SOC: soil organic carbon; crop PI: crop productivity index; PPT: precipitation; TMP: temperature.

Table A2. Land use and land cover datasets used for comparison.

Data variables	Time period	Resolution	Data sources
ERS major land uses	1945–2012	State level 4- to 5-year interval	Major uses of land in the United States, 2012. https://www.ers.usda.gov/data-products/major-land-uses/ (last access: 10 July 2022)
Land Use Harmonization (LUH2)	1600–2020	0.25° Annual	https://luh.umd.edu/ (last access: 3 January 2022)
Cropland density (YLmap)	1850–2016	1 km Annual	Yu and Lu (2017) https://doi.pangaea.de/10.1594/PANGAEA.881801
Historical fractional cropland areas (ZCmap)	1850–2000	5 arcmin Annual	Zumkehr and Campbell (2013) https://portal.nersc.gov/project/m2319/ (last access: 29 July 2022)
Hay area	1840–2012	County level 10-year interval	Haines et al. (2018) https://www.icpsr.umich.edu/web/ICPSR/studies/35206# (last access: 29 July 2022)
Historical LULC dataset	1938–1992	250 m Annual	Sohl et al. (2016) https://www.sciencebase.gov/catalog/item/59d3c73de4b05fe04cc3d1d1 (last access: 29 July 2022)
Crop area	1840–2017	County-level 10-year interval	Crossley (2020) https://www.openicpsr.org/openicpsr/project/115795/version/V3/view (last access: 11 November 2022)

Note: ERS: Economic Research Service, U.S. Department of Agriculture; YLmap: cropland density (Yu and Lu, 2017); ZCmap: historical fractional cropland areas (Zumkehr and Campbell, 2013).

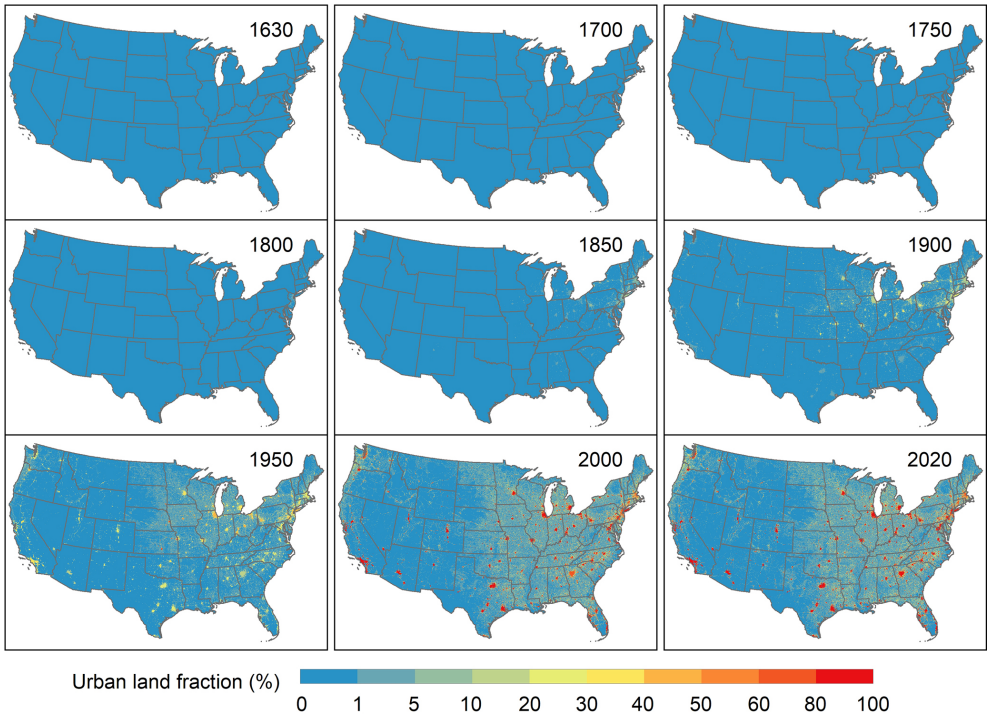

Figure A2. Fractional area of urban land in the conterminous United States during 1630–2020.

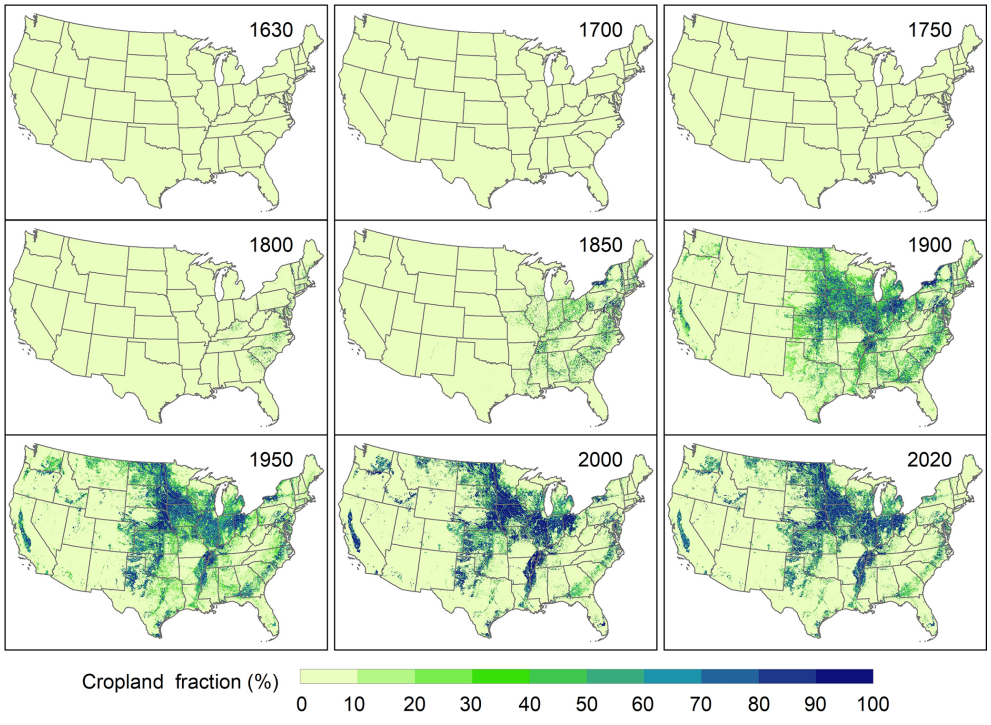

Figure A3. Fractional area of cropland in the conterminous United States during 1630–2020.

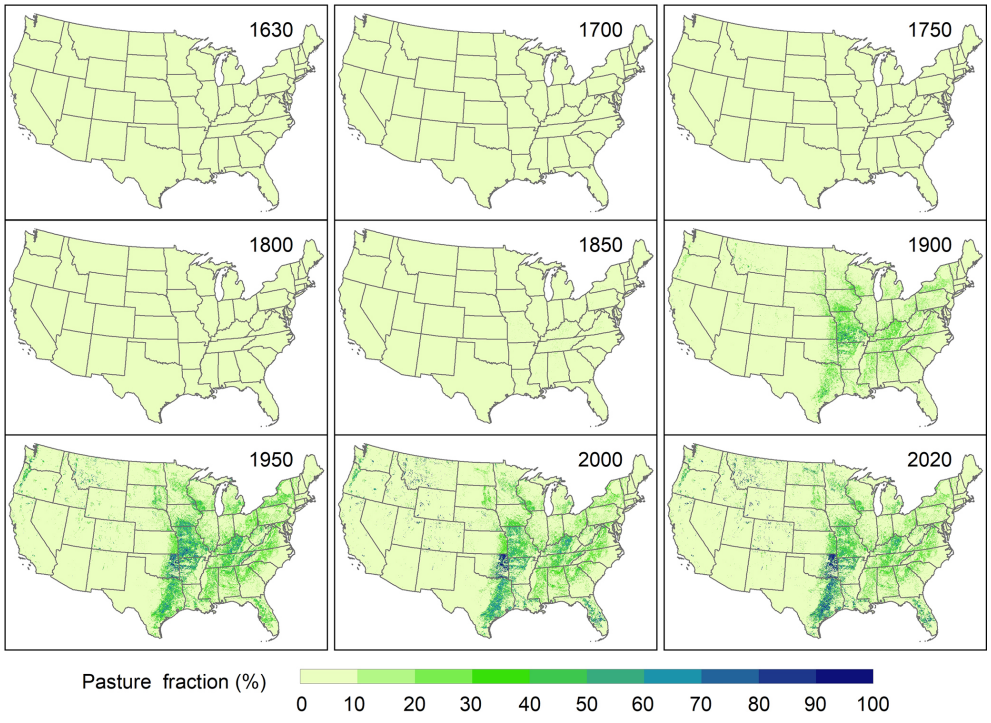

Figure A4. Fractional area of pasture in the conterminous United States during 1630–2020.

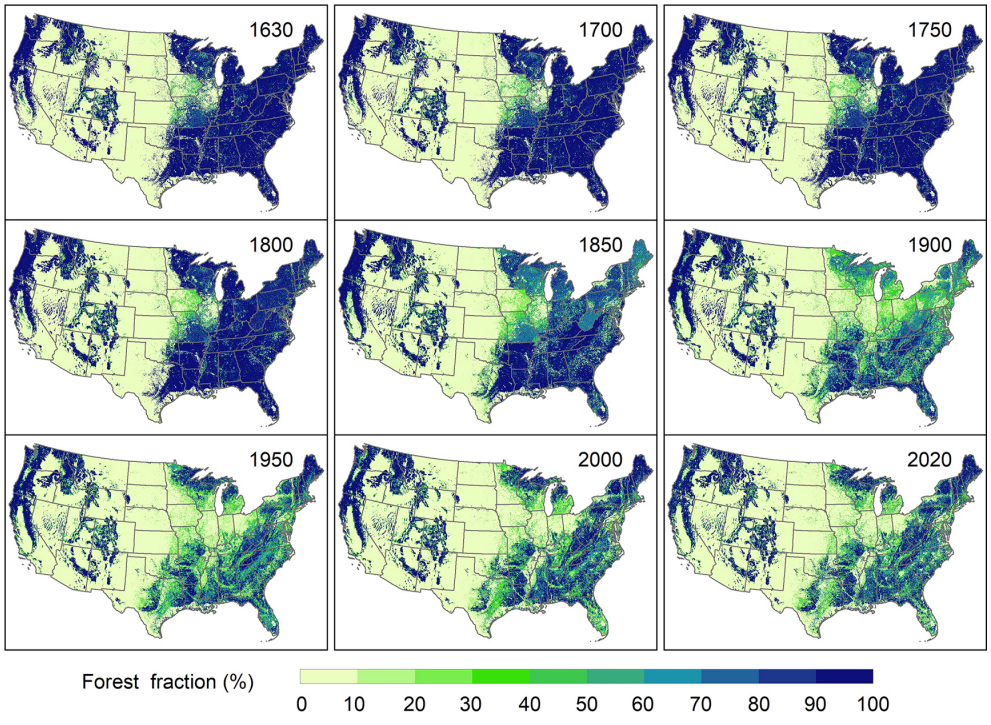

Figure A5. Fractional area of forest in the conterminous United States during 1630–2020.

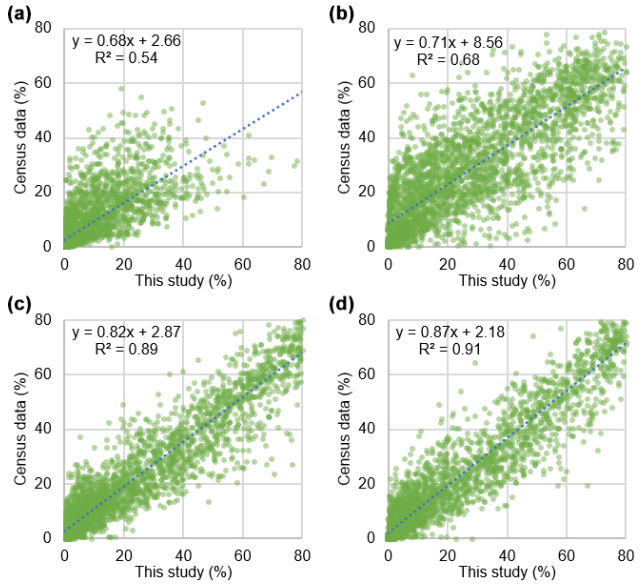

Figure A6. Statistical comparison between cropland area from this study and census data at county level in 1850 **(a)**, 1920 **(b)**, 1959 **(c)**, and 2002 **(d)**.

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/essd-15-1005-2023-supplement.

Author contributions. HT designed the research; XL implemented the research and analyzed the results; XL, HT, SP, and CL wrote and revised the paper.

Competing interests. At least one of the (co-)authors is a member of the editorial board of *Earth System Science Data*. The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. We thank the anonymous reviewers for their very constructive comments that have helped us to significantly improve this work.

Financial support. This study has been supported in part by the National Science Foundation (grant nos. 1903722 and 1922687), the National Oceanic and Atmospheric Administration (grant nos. NA16NOS4780204 and NA16NOS4780207), the National Aeronautics and Space Administration (grants nos. NNX12AP84G, NNX14AO73G, and NNX10AU06G) and the Department of the Treasury in cooperation with the State of Alabama Department of Conservation and Natural Resources (grant no. DISL-MESC-ALCOE-06).

Review statement. This paper was edited by Christopher Neigh and reviewed by four anonymous referees.

References

Bartholome, E. and Belward, A. S.: GLC2000: a new approach to global land cover mapping from Earth observation data, Int. J. Remote Sens., 26, 1959–1977, https://doi.org/10.1080/01431160412331291297, 2005.

Bigelow, D. P. and Borchers, A.: Major Uses of Land in the United States 2012, U.S. Department of Agriculture, Economic Research Service, https://www.ers.usda.gov/publications/pub-details/?pubid=84879 (last access: 13 February 2023), 2017.

Billington, R. A. and Ridge, M.: Westward expansion: a history of the American frontier, University of New Mexico Press, ISBN 9780826319814, 2001.

Boryan, C., Yang, Z., Mueller, R., and Craig, M.: Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program, Geocarto. Int., 26, 341–358, https://doi.org/10.1080/10106049.2011.562309, 2011.

Cao, B., Yu, L., Li, X., Chen, M., Li, X., Hao, P., and Gong, P.: A 1 km global cropland dataset from 10 000 BCE to 2100 CE, Earth Syst. Sci. Data, 13, 5403–5421, https://doi.org/10.5194/essd-13-5403-2021, 2021.

Chen, G., Pan, S., Hayes, D. J., and Tian, H.: Spatial and temporal patterns of plantation forests in the United States since the 1930s: an annual and gridded data set for regional Earth system modeling, Earth Syst. Sci. Data, 9, 545–556, https://doi.org/10.5194/essd-9-545-2017, 2017.

Chen, H., Tian, H., Liu, M., Melillo, J., Pan, S., and Zhang, C.: Effect of Land-Cover Change on Terrestrial Carbon Dynamics in the Southern United States, J. Environ. Qual., 35, 1533–1547, https://doi.org/10.2134/jeq2005.0198, 2006.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., and Mills, J.: Global land cover mapping at 30 m resolution: A POK-based operational approach, ISPRS. J. Photogramm. Remote Sens., 103, 7–27, https://doi.org/10.1016/j.isprsjprs.2014.09.002, 2015.

Clawson, M.: Forests in the long sweep of American history, Science, 204, 1168–1174, https://doi.org/10.1126/science.204.4398.1168, 1979.

Cole, K. L., Davis, M. B., Stearns, F., Guntenspergen, G., and Walker, K.: Historical landcover changes in the Great Lakes region, U.S. Geological Survey, Biological Resources Division, https://hdl.handle.net/11299/165997 (last access: 13 February 2023), 1998.

Coulson, D. P. and Joyce, L.: United States state-level population estimates: Colonization to 1999, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, https://doi.org/10.2737/RMRS-GTR-111, 2003.

Crossley, M. S.: County-level crop area in the USA 1840–2017, Inter-university Consortium for Political and Social Research, Ann Arbor, MI, https://doi.org/10.3886/E115795V3, 2020.

Crossley, M. S., Burke, K. D., Schoville, S. D., and Radeloff, V. C.: Recent collapse of crop belts and declining diversity of US agriculture since 1840, Glob. Change Biol., 27, 151–164, https://doi.org/10.1111/gcb.15396, 2021.

Dangal, S. R. S., Felzer, B. S., and Hurteau, M. D.: Effects of agriculture and timber harvest on carbon sequestration in the eastern US forests, J. Geophys. Res.-Biogeo., 119, 35–54, https://doi.org/10.1002/2013JG002409, 2014.

Domke, G. M., Oswalt, S. N., Walters, B. F., and Morin, R. S.: Tree planting has the potential to increase carbon sequestration capacity of forests in the United States, P. Natl. Acad. Sci. USA, 117, 24649–24651, https://doi.org/10.1073/pnas.2010840117, 2020.

Drummond, M. A. and Loveland, T. R.: Land-use pressure and a transition to forest-cover loss in the eastern United States, BioScience, 60, 286–298, https://doi.org/10.1525/bio.2010.60.4.7, 2010.

Ellis, E. C., Gauthier, N., Klein Goldewijk, K., Bliege Bird, R., Boivin, N., Díaz, S., Fuller, D. Q., Grill J. L., Kaplan, J. O., Kingston, N., Locke, H., McMichael, C. N. H., Ranco, D., Rick, T. C., Shaw, R. M., Stephens, L., Svenning, J., and Watson, J. E. M.: People have shaped most of terrestrial nature for at least 12,000 years, P. Natl. Acad. Sci. USA, 118, e2023483118, https://doi.org/10.1073/pnas.2023483118, 2021.

Fang, Y. and Jawitz, J. W.: High-resolution reconstruction of the United States human population distribution, 1790 to 2010, Sci. Data, 5, 180067, https://doi.org/10.1038/sdata.2018.67, 2018.

Foster, D. R.: Land-Use History (1730–1990) and Vegetation Dynamics in Central New-England, USA, J. Ecol., 80, 753–772, https://doi.org/10.2307/2260864, 1992.

Foster, D. R., Motzkin, G., and Slater, B.: Land-use history as long-term broad-scale disturbance: regional forest dynamics in central New England, Ecosystems, 1, 96–119, https://doi.org/10.1007/s100219900008, 1998.

Fretwell, J. D., Williams, J. S., and Redman, P. J.: National water summary on wetland resources, U.S. Government Printing Office, https://doi.org/10.3133/wsp2425, 1996.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X.: MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, Remote Sens. Environ., 114, 168–182, https://doi.org/10.1016/j.rse.2009.08.016, 2010.

Fuchs, R., Herold, M., Verburg, P. H., and Clevers, J. G. P. W.: A high-resolution and harmonized model approach for reconstructing and analysing historic land changes in Europe, Biogeosciences, 10, 1543–1559, https://doi.org/10.5194/bg-10-1543-2013, 2013.

Garrison, C. E.: Forestry and Tree Planting in Virginia, Reforestation, Nurseries, and Genetic Resources (RNGR), https://rngr.net/publications/tpn/55-2/forestry-and-tree-planting-in-virginia (last access: 13 February 2023), 2012.

Grassi, G., House, J., Dentener, F., Federici, S., den Elzen, M., and Penman, J.: The key role of forests in meeting climate targets requires science for credible mitigation, Nature Clim. Change, 7, 220–226, https://doi.org/10.1038/nclimate3227, 2017.

Griscom, B. W., Adams, J., Ellis, P. W., Houghton, R. A., Lomax, G., Miteva, D. A., Schlesinger, W. H., Shoch, D., Siikamaki, J. V., Smith, P., Woodbury, P., Zganjar, C., Blackman, A., Campari, J., Conant, R. T., Delgado, C., Elias, P., Gopalakrishna, T., Hamsik, M. R., Herrero, M., Kiesecker, J., Landis, E., Laestadius, L., Leavitt, S. M., Minnemeyer, S., Polasky, S., Potapov, P., Putz, F. E., Sanderman, J., Silvius, M., Wollenberg, E., and Fargione, J.: Natural climate solutions, P. Natl. Acad. Sci. USA, 114, 11645–11650, https://doi.org/10.1073/pnas.1710465114, 2017.

Hagenauer, J. and Helbich, M.: A geographically weighted artificial neural network, Int. J. Geogr. Inform. Sci., 36, 215–235, https://doi.org/10.1080/13658816.2021.1871618, 2022.

Haines, M., Fishback, P., and Rhode, P.: United States Agriculture Data, 1840–2012, Inter-university Consortium for Political and Social Research, https://doi.org/10.3886/ICPSR35206.v4, 2018.

Hall, B., Motzkin, G., Foster, D. R., Syfert, M., and Burk, J.: Three hundred years of forest and land-use change in Massachusetts, USA, J. Biogeogr., 29, 1319–1335, https://doi.org/10.1046/j.1365-2699.2002.00790.x, 2002.

Hanberry, B. B., Kabrick, J. M., He, H. S., and Palik, B. J.: Historical trajectories and restoration strategies for the Mississippi River Alluvial Valley, For. Eco. Manag., 280, 103–111, https://doi.org/10.1016/j.foreco.2012.05.033, 2012.

Heimlich, R. E. and Daugherty, A. B.: America's cropland: Where does it come from, United States Department of Agriculture, 3–9, https://handle.nal.usda.gov/10113/IND20394000 (last access: 13 February 2023), 1991.

Hart, J. F.: Loss and abandonment of cleared farm land in the Eastern United States, An. Assoc. Am. Geogr., 58, 417–440, 1968.

He, F., Li, S., and Zhang, X.: A spatially explicit reconstruction of forest cover in China over 1700–2000, Glob. Planet. Change, 131, 73–81, https://doi.org/10.1016/j.gloplacha.2015.05.008, 2015.

Homer, C., Dewitz, J., Jin, S., Xian, G., Costello, C., Danielson, P., Gass, L., Funk, M., Wickham, J., Stehman, S., Auch, R., and Riitters, K.: Conterminous United States land cover change patterns 2001–2016 from the 2016 National Land Cover Database, ISPRS. J. Photogramm. Remote, 162, 184–199, https://doi.org/10.1016/j.isprsjprs.2020.02.019, 2020.

Houghton, R. A., Hackler, J. L., and Lawrence, K. T.: The US carbon budget: Contributions from land-use change, Science, 285, 574–578, https://doi.org/10.1126/science.285.5427.574, 1999.

Hurt, R. D.: American agriculture: A brief history, Purdue University Press, 2002.

Hurtt, G. C., Frolking, S., Fearon, M. G., Moore, B., Shevliakova, E., Malyshev, S., Pacala, S. W., and Houghton, R. A.: The underpinnings of land-use history: three centuries of global gridded land-use transitions, wood-harvest activity, and resulting secondary lands, Glob. Change Biol., 12, 1208–1229, https://doi.org/10.1111/j.1365-2486.2006.01150.x, 2006.

Hurtt, G. C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J. C., Fisk, J., Fujimori, S., Klein Goldewijk, K., Hasegawa, T., Havlik, P., Heinimann, A., Humpenöder, F., Jungclaus, J., Kaplan, J. O., Kennedy, J., Krisztin, T., Lawrence, D., Lawrence, P., Ma, L., Mertz, O., Pongratz, J., Popp, A., Poulter, B., Riahi, K., Shevliakova, E., Stehfest, E., Thornton, P., Tubiello, F. N., van Vuuren, D. P., and Zhang, X.: Harmonization of global land use change and management for the period 850–2100 (LUH2) for CMIP6, Geosci. Model Dev., 13, 5425–5464, https://doi.org/10.5194/gmd-13-5425-2020, 2020.

Jeon, S. B., Olofsson, P., and Woodcock, C. E.: Land use change in New England: a reversal of the forest transition, J. Land Use Sci.,

9, 105–130, https://doi.org/10.1080/1747423X.2012.754962, 2014.

Klein Goldewijk, K., Beusen, A., Doelman, J., and Stehfest, E.: Anthropogenic land use estimates for the Holocene – HYDE 3.2, Earth Syst. Sci. Data, 9, 927–953, https://doi.org/10.5194/essd-9-927-2017, 2017.

Lark, T. J., Mueller, R. M., Johnson, D. M., and Gibbs, H. K.: Measuring land-use and land-cover change using the US department of agriculture's cropland data layer: Cautions and recommendations, Int. J. Appl. Earth Obs. Geoinf., 62, 224–235, https://doi.org/10.1016/j.jag.2017.06.007, 2017.

Lark, T. J., Spawn, S. A., Bougie, M., and Gibbs, H. K.: Cropland expansion in the United States produces marginal yields at high costs to wildlife, Nat. Commun., 11, 1–11, https://doi.org/10.1038/s41467-020-18045-z, 2020.

Lark, T. J., Schelly, I. H., and Gibbs, H. K.: Accuracy, Bias, and Improvements in Mapping Crops and Cropland across the United States Using the USDA Cropland Data Layer, Remote Sens., 13, 968, https://doi.org/10.3390/rs13050968, 2021.

Leyk, S. and Uhl, J. H.: HISDAC-US, historical settlement data compilation for the conterminous United States over 200 years, Sci. Data, 5, 1–14, https://doi.org/10.1038/sdata.2018.175, 2018.

Leyk, S., Uhl, J. H., Connor, D. S., Braswell, A. E., Mietkiewicz, N., Balch, J. K., and Gutmann, M.: Two centuries of settlement and urban development in the United States, Sci. Adv., 6, eaba2937, https://doi.org/10.1126/sciadv.aba2937, 2020.

Li, S., He, F., and Zhang, X.: A spatially explicit reconstruction of cropland cover in China from 1661 to 1996, Regional Environmental Change, 16, 417–428, https://doi.org/10.1007/s10113-014-0751-4, 2016.

Li, X., Yu, L., Sohl, T., Clinton, N., Li, W., Zhu, Z., Liu, X., and Gong, P.: A cellular automata downscaling based 1 km global land use datasets (2010–2100), Sci. Bull., 61, 1651–1661, https://doi.org/10.1007/s11434-016-1148-1, 2016.

Li, X., Tian, H., Pan, S., and Lu, C.: Land use and land cover changes in the contiguous United States from 1630 to 2020 (v2.0), Zenodo [data set], https://doi.org/10.5281/zenodo.7055086, 2022.

Liu, M. and Tian, H.: China's land cover and land use change from 1700 to 2005: Estimations from high-resolution satellite data and historical archives, Global Biogeochem. Cy., 24, GB3003, https://doi.org/10.1029/2009gb003687, 2010.

Liu, X., Liang, X., Li, X., Xu, X., Ou, J., Chen, Y., Li, S., Wang, S., and Pei, F.: A future land use simulation model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects, Landsc. Urban Plan., 168, 94–116, https://doi.org/10.1016/j.landurbplan.2017.09.019, 2017.

MacCleery, D. W.: American forests: a history of resiliency and recovery, U.S. Department of Agriculture, Forest History Society, ISBN 0890300488, 2011.

Meinig, D. W.: The Shaping of America: A Geographical Perspective on 500 Years of History, vol. 2: Continental America, 1800–1867, Yale University Press, https://www.jstor.org/stable/j.ctt5hk0p1 (last access: 13 February 2023), 1993.

Mergener, R., Botti, W., and Heyd, R.: Forestry and Tree Planting in Michigan, Reforestation, Nurseries, and Genetic Resources (RNGR), https://rngr.net/publications/tpn/57-1/forestry-and-tree-planting-in-michigan (last access: 13 February 2023), 2014.

Olmstead, A. L. and Rhode, P. W.: A history of California agriculture, Giannini Foundation of Agricultural Economics, University of California, https://giannini.ucop.edu/ (last access: 13 February 2023), 2017.

Oswalt, S. N., Smith, W. B., Miles, P. D. and Pugh, Scott, A.: Forest Resources of the United States, 2012: a technical document supporting the Forest Service 2010 update of the RPA Assessment, U.S. Department of Agriculture, Forest Service, Washington Office, https://doi.org/10.2737/WO-GTR-91, 2014.

Oswalt, S. N., Miles, P. D., Pugh, S. A., and Smith, W. B.: Forest Resources of the United States, 2017: a technical document supporting the Forest Service 2020 RPA Assessment, U.S. Department of Agriculture, Forest Service, Washington Office, https://doi.org/10.2737/WO-GTR-97, 2019.

Peng, S., Ciais, P., Maignan, F., Li, W., Chang, J., Wang, T., and Yue, C.: Sensitivity of land use change emission estimates to historical land use and land cover mapping, Global Biogeochem. Cy., 31, 626–643, https://doi.org/10.1002/2015gb005360, 2017.

Reuss, L. A., Wooten, H. H., and Marschner, F. J.: Inventory of Major Land Uses in the United States, U.S. Department of Agriculture, https://ageconsearch.umn.edu/record/314797 (last access: 13 February 2023), 1948.

Rollins, M. G.: LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment, Int. J. Wildland Fire, 18, 235–249, https://doi.org/10.1071/WF08088, 2009.

Schulman, S. A.: The Lumber Industry of the Upper Cumberland River Valley, Tennessee Historical Quarterly, 32, 255–264, https://www.jstor.org/stable/42623392 (last access: 3 February 2023), 1973.

Smith, W. B., Vissage, J. S., Darr, D. R., and Sheffield, R. M.: Forest Resources of the United States, 1997, U.S. Department of Agriculture Forest Service, https://doi.org/10.2737/NC-GTR-219, 2001.

Sohl, T., Reker, R., Bouchard, M., Sayler, K., Dornbierer, J., Wika, S., Quenzer, R., and Friesz, A.: Modeled historical land use and land cover for the conterminous United States, J. Land Use Sci., 11, 476–499, https://doi.org/10.1080/1747423X.2016.1147619, 2016.

Sohl, T. L., Sayler, K. L., Bouchard, M. A., Reker, R. R., Friesz, A. M., Bennett, S. L., Sleeter, B. M., Sleeter, R. R., Wilson, T., and Soulard, C.: Spatially explicit modeling of 1992–2100 land cover and forest stand age for the conterminous United States, Ecol. Appl., 24, 1015–1036, https://doi.org/10.1890/13-1245.1, 2014.

Stanturf, J. A., Palik, B. J., and Dumroese, R. K.: Contemporary forest restoration: A review emphasizing function, For. Ecol. Manag., 331, 292–323, https://doi.org/10.1016/j.foreco.2014.07.029, 2014.

Steyaert, L. T. and Knox, R. G.: Reconstructed historical land cover and biophysical parameters for studies of land-atmosphere interactions within the eastern United States, J. Geophys. Res.-Atmos., 113, D02101, https://doi.org/10.1029/2006jd008277, 2008.

Thompson, J. R., Carpenter, D. N., Cogbill, C. V., and Foster, D. R.: Four Centuries of Change in Northeastern United States Forests, Plos One, 8, e72540, https://doi.org/10.1371/journal.pone.0072540, 2013.

Tian, H., Chen, G., Zhang, C., Liu, M., Sun, G., Chappelka, A., Ren, W., Xu, X., Lu, C., and Pan, S.: Century-scale responses

of ecosystem carbon storage and flux to multiple environmental changes in the southern United States, Ecosystems, 15, 674–694, https://doi.org/10.1007/s10021-012-9539-x, 2012.

Tian, H., Banger, K., Tao, B., and Dadhwal, V. K.: History of land use in India during 1880–2010: Large-scale land transformation reconstructed from satellite data and historical achieves, Glob. Planet. Change, 121, 76–88, https://doi.org/10.1016/j.gloplacha.2014.07.005, 2014.

Tian, H., Xu, R., Pan, S., Yao, Y., Bian, Z., Cai, W., Hopkinson, C., Justic, D., Lohrenz, S., Lu, C., Ren, W., and Yang, J.: Long-term trajectory of nitrogen loading and delivery from Mississippi River Basin to the Gulf of Mexico, Global Biogeochem. Cy., 34, e2019GB006475, https://doi.org/10.1029/2019GB006475, 2020.

Uhl, J. H., Leyk, S., McShane, C. M., Braswell, A. E., Connor, D. S., and Balk, D.: Fine-grained, spatiotemporal datasets measuring 200 years of land development in the United States, Earth Syst. Sci. Data, 13, 119–153, https://doi.org/10.5194/essd-13-119-2021, 2021.

U.S. Department of Agriculture: Summary Report: 2017 National Resources Inventory, Natural Resources Conservation Service Washington, D.C., and Center for Survey Statistics and Methodology, Iowa State University, Ames, Iowa, https://www.nrcs.usda.gov/nri (last access: 13 February 2023), 2020.

U.S. Department of Agriculture and Economic Research Service: Our Land and Water Resources: Current and Prospective Supplies and Uses, U.S. Government Printing Office, https://www.govinfo.gov/app/details/CZIC-s21-a46-no-1290 (last access: 13 February 2023), 1974.

Verburg, P. H. and Overmars, K. P.: Combining top-down and bottom-up dynamics in land use modeling: exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model, Landscape Ecol., 24, 1167–1181, https://doi.org/10.1007/s10980-009-9355-7, 2009.

Verburg, P. H., Schulp, C. J. E., Witte, N., and Veldkamp, A.: Downscaling of land use change scenarios to assess the dynamics of European landscapes, Agr. Ecosyst. Environ., 114, 39–56, https://doi.org/10.1016/j.agee.2005.11.024, 2006.

Waisanen, P. J. and Bliss, N. B.: Changes in population and agricultural land in conterminous United States counties, 1790 to 1997, Global Biogeochem. Cy., 16, 84-81–84-19, https://doi.org/10.1029/2001gb001843, 2002.

West, T. O., Page, Y. L., Huang, M., Wolf, J., and Thomson, A. M.: Downscaling global land cover projections from an integrated assessment model for use in regional analyses: results and evaluation for the US from 2005 to 2095, Environ. Res. Lett., 9, 064004, https://doi.org/10.1088/1748-9326/9/6/064004, 2014.

Winkler, K., Fuchs, R., Rounsevell, M., and Herold, M.: Global land use changes are four times greater than previously estimated, Nat. Commun., 12, 1–10, https://doi.org/10.1038/s41467-021-22702-2, 2021.

Williams, M.: Americans and Their Forests: A Historical Geography, Cambridge University Press, ISBN 9780521428378, 1992.

Yang, J., Tao, B., Shi, H., Ouyang, Y., Pan, S., Ren, W., and Lu, C.: Integration of remote sensing, county-level census, and machine learning for century-long regional cropland distribution data reconstruction, Int. J. Appl. Earth Obs. Geoinf., 91, 102151, https://doi.org/10.1016/j.jag.2020.102151, 2020.

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S. M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G. C., Rigge, M., and Xian, G.: A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies, ISPRS. J. Photogramm. Remote, 146, 108–123, https://doi.org/10.1016/j.isprsjprs.2018.09.006, 2018.

Yu, Z. and Lu, C.: Historical cropland of the continental U.S. from 1850 to 2016, PANGAEA [data set], https://doi.org/10.1594/PANGAEA.881801, 2017.

Yu, Z. and Lu, C.: Historical cropland expansion and abandonment in the continental U.S. during 1850 to 2016, Glob. Ecol. Biogeogr., 27, 322–333, https://doi.org/10.1111/geb.12697, 2018.

Yu, Z., Lu, C., Tian, H., and Canadell, J. G.: Largely underestimated carbon emission from land use and land cover change in the conterminous United States, Glob. Change Biol., 25, 3741–3752, https://doi.org/10.1111/gcb.14768, 2019.

Zumkehr, A. and Campbell, J. E.: Historical U.S. Cropland areas and the potential for bioenergy production on abandoned croplands, Environ. Sci. Technol., 47, 3840–3847, https://doi.org/10.1021/es3033132, 2013.